COMPARATIVE ANALYSIS OF APOPTOTIC FUNCTION BETWEEN HUMANS, CHIMPANZEES AND MACAQUES

A Dissertation
Presented to
The Academic Faculty

by

Gaurav Arora

In Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy in the
School of Biology

Georgia Institute of Technology
August 2011

COMPARATIVE ANALYSIS OF APOPTOTIC FUNCTION BETWEEN HUMANS, CHIMPANZEES AND MACAQUES

Approved by:

Dr. John F. McDonald, Advisor
School of Biology
Georgia Institute of Technology

Dr. James Thomas
Department of Human Genetics
Emory University

Dr. I. King Jordan
School of Biology
Georgia Institute of Technology

Dr. Todd M. Preuss
Division of Neuroscience
Yerkes National Primate Research Center

Dr. Francesca Storici
School of Biology
Georgia Institute of Technology

Date Approved: June 23, 2011

For my Family and Friends

ACKNOWLEDGEMENTS

I would like to thank many people without whose support this dissertation would not have been possible. First and most important, I would like to thank my advisor, Dr. John McDonald, without whose encouragement I wouldn't have been able to complete five very important years of my professional life. I have learnt a lot under his guidance and will always be thankful for it.

I would like to thank my committee – Dr. I.King Jordan, Dr. Francesca Storici, Dr. James Thomas and Dr. Todd Preuss – for their insights, knowledge and suggestions throughout my graduate years. Their inputs have really made me more appreciative of the scientific method and have given me a much better understanding of my field.

Special thanks to Roman Mezencev, without whose help and knowledge this dissertation would not have happened. Roman has been a great driving force in my project, and I will always be thankful to him. Also a special thanks to Nathan Bowen for all his advice and help rendered during my formative years at Georgia Tech.

I would like to thank each and every member of the McDonald Lab for all the support, understanding and fun throughout my years in the lab. Thanks, DeEtte Walker, Lilya Matyunina, Kenneth Scarberry, Lijuan Wang, Shubin Shahab, Andrew Huang, Vinay Mittal, Chris Hill, Loukia Lili and Neda Jabbari. Further thanks to past members of the McDonald Lab, especially Erin Dickerson, Laura Kapa, Jing Chen, Nina Schubert,

Marine Cazuax and Travis Wagner. A special thanks to a great friend, Nalini Polavarapu for her support, help and collaboration in my projects. I couldn't have asked for a better lab to work in at Georgia Tech. Thank You.

I am very grateful to a number of people at Georgia Tech, who made sure that I moved forward throughout my years at Tech. Special thanks to Kevin Roman for all his help and guidance during the last few years. Also thanks to Lisa Tuttle for her help in the School of Biology office. A special thanks to Cara Gormally for all her help and guidance during my teaching years at Tech.

I am very grateful to all the wonderful friends I have met at Tech. Friends who have been a part of every celebration during our graduate years. Thank you, Lee Katz, Samantha Stuckey, Ahsan Huda, Patrick Ruff, Navin Elango, Harjeet Johal and Carolee Shackelford.

I am also grateful to all my friends at Georgia State University in downtown Atlanta. These are friends with whom I have grown, after I first landed in Atlanta from India. Thank you, Miki Kassai, Jeanetta Floyd, Subrata Mishra, Rachael Farah-Abraham, Aga Truax, Jaya Punetha, Abhishek Gupta, Judy Block, Deidre Steed, Martha Fowler and Donnia Turner.

A special thanks to my very good friend, Patra Volarath, for standing by me through thick and thin, through day and night. Without her encouragement, I wouldn't have been able to go through gloomy and happy days.

A big thank you to all my friends from India for believing in me and checking on me periodically. Thank you for your support, Meera Simha, Stuti Prasad, Menaz Bhanpurwala, Tasneem Khan, Tina Pithawala, Lochan Kothari, Feroz Merchant, Kirtan Koticha and Saleel Raut.

I also want to thank my cousins, Ritu Chandok, Rajeev Chandok and Geeta Arora, for supporting my decisions and adding their inputs to help me make better choices. Special thanks to my aunts, Mrs. Josephine Arora, Mrs. Shashi Arora, Mrs.Vinod Grover and Mrs. Jyoti Punjabi for all their love and support.

I would like to thank my partner, Gaurav Mathur, for his relentless patience, love and support. Without him, I wouldn't have been able to reach this important milestone in my life.

Last but not least, I am grateful to my family for their undying love and support. Thank you, my brother Mohit Arora and sister-in-law Ritu Taimni-Arora, for always being there for me.

Finally I come to the most beautiful woman in the world, my mom, Deepa Arora. Without her love, patience and support, none of this would have been possible. My father, the late Sureshkumar Arora, who is always in my memory, constantly supported my decisions and career path. I dedicate this dissertation to both of my parents.

Thank You.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF SYMBOLS AND ABBREVIATIONS

↑	Deletion of sequence
↓	Insertion of sequence
$\Delta\Psi_m$	Mitochondrial Transmembrane Electrical Potential Gradient
*	underestimate of INDEL variation in upstream region due to poor sequence quality
A	Atlas of Genetics and Cytogenetics in Oncology and Haematology
Ad	Adenine
AIDS	Acquired Immunodeficiency Syndrome
ANOVA	Analysis of Variance
Asp	Aspartate
Arg	Arginine
bps	Base Pairs
C	Chimpanzee
CDs	Chimpanzee Deletions
CDS	Coding Sequences
CERV	Chimpanzee Endogenous Retrovirus
CGs	Chimpanzee Gaps
CIs	Chimpanzee Insertions
CpG	Cytosine Phosphate Guanine region
Cy	Cytosine
Cys	Cysteine
DD	Death Domain
DE	Differentially Expressed

DED	Death Effector Domain
DISC	Death Inducing cell Signaling Complex
DMEM	Dulbecco's Modified Eagle Medium
DNA	Deoxyribonucleic Acid
ENSEMBL	European Bioinformatics Institute database
Exp	Detected but not differentially Expressed
ERVs	Endogenous Retroviruses
FasL	Fas Ligand
FasR	Fas Receptor
FADD	Fas-Associated protein with Death Domain
FBS	Fetal Bovine Serum
Gu	Guanine
GC	Guanine-cytosine content
GenMAPP	Gene Map Annotator and Pathway Profiler
GSEA	Gene Set Enrichment Analysis
H	Human
HIV	Human Immunodeficiency Virus
HDs	Human Deletions
hg16	Human Genome Version 16
hg18	Human Genome Version 18
hgU133plus2	Affymetrix Human Genome U133 Plus 2.0 Array Annotation Data
HGs	Human Gaps
HIs	Human Insertions
I	Inducer of Apoptosis
IC_{50}	Half maximal Inhibitory Concentration

INDEL	INsertion/DELetion
kb	kilobases
LINEs	Long Interspersed Nuclear Elements
LTR	Long Terminal Repeat
MAS	Microarray Suite
Mb	Megabases
MEs	Mosaic Elements
MMC	Mitomycin-C
MPT	Mitochondrial Permeability Transition
MYA	Million Years Ago
NISs	Non-Interspersed Sequences
panTro	*Pan troglodytes* (chimpanzee) genome
PBS	Phosphate Buffered Saline
PCD	Programmed Cell Death
Pro	Proline
PTERV	*Pan troglodytes* Endogenous Retrovirus
P(#)	Passage Number of Cell Lines
qrt-PCR	Quantitative Reverse Transcriptase Polymerase Chain Reaction
R	Repressor of Apoptosis
REs	Retrotransposon Sequences
rheMac	*Rhesus macaque* (macaque) genome
RNA	Ribonucleic Acid
RPMI	basal medium consisting of vitamins, amino acids, etc.
S	Ingenuity® System database
SINEs	Short Interspersed Nuclear Elements

SNPs	Single Nucleotide Polymorphisms
SVAs	composed of SINE, VNTRs and Alu elements
Th	Thymine
TEs	Transposable Elements
TFBS	Transcription Factor Binding Site
TMRE	Tetramethyl Rhodamine Ethyl
TMRM	Tetramethyl Rhodamine Methyl
TNF	Tumor Necrosis Factor
UCSC	University of California Santa Cruz
VNTRs	Variable Number of Tandem Repeats

SUMMARY

Despite the high degree of similarity between the human and chimpanzee genomes (98.5%), a number of phenotypic traits distinguish the two species. Some of these traits include a larger brain, with certain areas of specialized function, and a higher incidence of cancer in humans. One pathway that is common to both brain development and disease progression to cancer is apoptosis.

During neuronal development, apoptosis functions by controlling the number of cells. In cancer, however, the ability of the apoptotic pathway to limit uncontrolled cell growth is lost. This research aims at exploring whether there is a difference in the apoptotic turnover between the two species. If so, this difference may explain differences in the above phenotypic traits. This research also explores the genetic basis of the some of the phenotypic differences between the two species.

Microarray analysis, pathway prediction analysis, high throughput genome processing, statistical analysis, data management and experimental evidence were all used to achieve two major research advances in the comparison of humans and chimpanzees with respect to apoptotic function and in the understanding of the genetic basis of the phenotypic differences between the two species.

Research Advance 1: Gene expression analysis and pathway analysis were used to identify pathways that were enriched with genes that were differentially expressed between humans and chimpanzees. Based on these analyses, it was predicted that the

apoptotic pathway has a reduced function in humans compared to chimpanzees. Tests were then carried out on human, chimpanzee and macaque cell lines to compare the apoptotic turnover among these species. Comparing results from the tests showed that apoptosis has a reduced turnover in humans, compared to both chimpanzee and macaques. This reduced turnover may account for the larger size of the human brain (due to apoptotic turnover being reduced during development) and an increased susceptibility to cancer in humans (due to apoptosis being suppressed during cancer).

Research Advance 2: Microarray analysis, statistical analysis and high throughput genome processing was used to identify and associate INDEL variation (80-12,000 bp) with differential gene expression between humans and chimpanzees. A significant association was found between differential gene expression and INDEL variation between humans and chimpanzees, primarily involving the insertion of interspersed (predominantly retrotransposon) and non-interspersed sequences in the human lineage. The majority of this functionally significant INDEL variation was mapped to the introns and arose in the human lineage after humans and chimpanzees diverged from a common ancestor.

CHAPTER 1

INTRODUCTION AND LITERATURE SURVEY

Humans (*Homo sapiens)* differ from their closest primate relatives in a number of characteristics. Some of the characteristics that are human-specific include bipedalism, an exceptionally large brain, use of complex language and other unusual cognitive and behavioral abilities (Gagneux and Varki 2001; Hacia 2001; Carroll 2003; de Sousa and Wood 2007; Varki 2007). The disease profile of humans differs from that of other primates, e.g. humans are vulnerable to neurodegenerative diseases like Huntington's and Alzheimer's (Gearing et al. 1994; Poduri et al. 1994; Walker 1999). Additionally, humans not only show a progression from HIV to AIDS (Rutjens et al. 2003), but also show a higher susceptibility to certain cancers (McClure 1973). The high susceptibility to neurodegenerative diseases and cancer in humans may be linked to the fact that among all the primates, humans have the longest lifespan (Hawkes et al. 1998; Blurton Jones et al. 2002).

Much before the era of modern biology, Huxley (Huxley 1863) and Darwin (Darwin 1871) postulated that among the primates, the African great apes, i.e. chimpanzees *(Pan troglodytes)* and gorillas *(Gorilla gorilla)*, are the closest relatives of the humans. DNA sequence analysis and cytogenetic studies have since shown that the genomes of humans are strikingly similar to the genomes of the African great apes (Sibley and Ahlquist 1987; Caccone and Powell 1989). Independent studies done on chimpanzees have shown that they also engage in cultural practices (Whiten et al. 1999), use tools (Biro et al. 2003;

Yamakoshi and Myowa-Yamakoshi 2004) and display rudimentary forms of language (Savage-Rumbaugh et al. 1978; Savage-Rumbaugh et al. 1985; Bodamer and Gardner 2002). It has been proposed that focusing on the differences between the genomes of humans and the African great apes would provide a better understanding of the genetic basis for distinctly human traits (Varki 2000; Gagneux and Varki 2001).

This chapter explores the known genetic differences between humans and African great apes, especially the chimpanzees. A major gap in the previous studies is the lack of data on the differences between humans and chimpanzees in pathway function. Pathways like apoptosis are important during development of human specific characteristics like a bigger brain and the susceptibility of humans to diseases like cancer. This chapter reviews the mechanism of apoptosis in detail and explores its role in the known phenotypic differences between humans and chimpanzees. Later on in the chapter, the potential role of insertion/deletion (INDELs) events to the phenotypic differences between humans and chimpanzees is explored.

Among the African great apes, the chimpanzees are the closest relatives of the humans

The first studies (Sibley and Ahlquist 1987; Caccone and Powell 1989) comparing the large-scale DNA sequences between human and African great ape genomes reported estimates of divergence dates between humans and chimpanzee lineages as 5.8-7.1 million years ago (MYA) and between human and gorilla lineages as 8.3-10.1 MYA. These studies used hybridization melting curve analysis of non-repetitive human and

great ape DNA sequences and showed that the human and chimpanzee non-repetitive DNA sequence had more closely related melting points than the human and gorilla sequences.

A number of comparative studies done on a large number of human, chimpanzee and gorilla sequenced loci, as well as one phylogenetic study (Figure 1.1) (Goodman 1996), support the human-chimpanzee dichotomy (Ruvolo 1997; Chen and Li 2001; Kaessmann et al. 2001) rather than a human-chimpanzee-gorilla trichotomy (Satta et al. 2000). The main goal of one of these comparative sequence studies (Chen and Li 2001) was to determine the sequence similarities between the human, chimpanzee and gorilla genomes. Sequence comparisons were done on randomly distributed 24 kb intergenic non-repetitive sequences, and it was shown that the human sequences were 98.67% similar to the chimpanzee sequences and 98.38% similar to the gorilla sequences. In this study (Chen and Li 2001), it was also shown that the common ancestor of the human and chimpanzees existed 4.6-6.2 MYA, and that the common ancestor of the humans and gorillas existed 6.3-8.5 MYA.

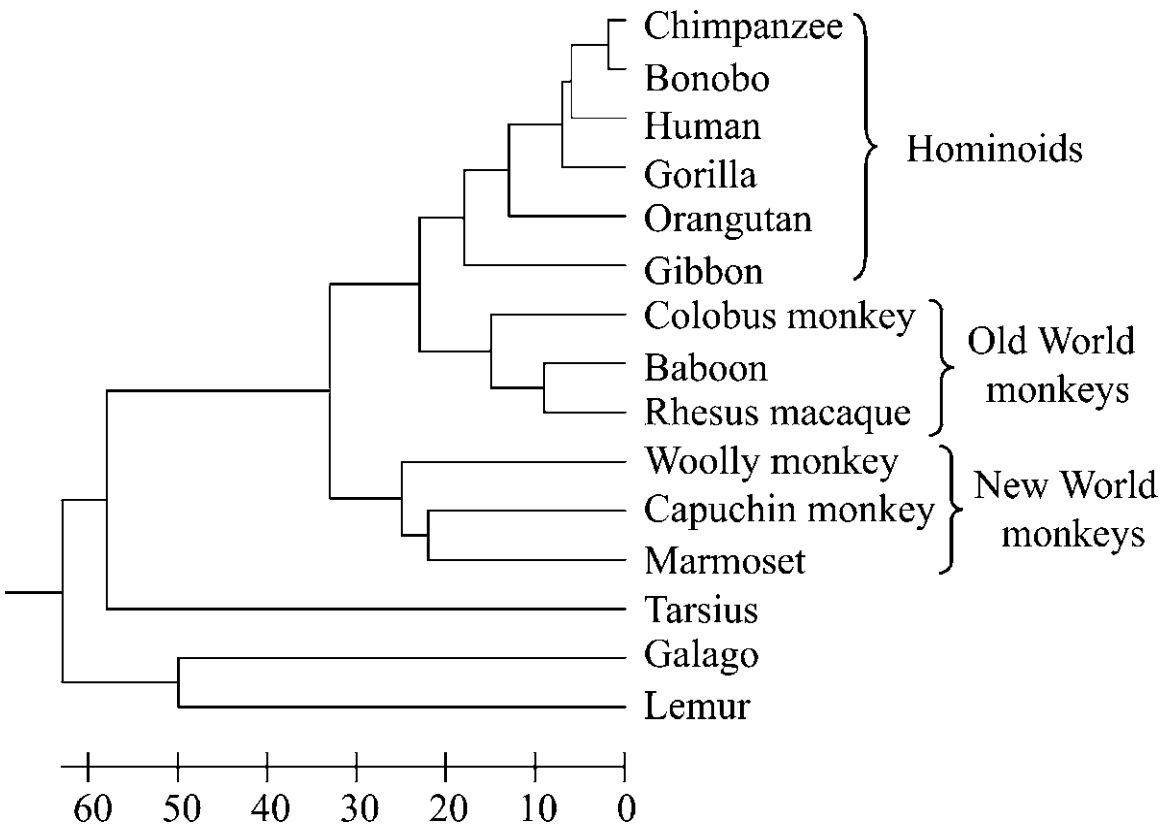


Figure 1.1: Phylogenetic comparisons of primate species. Scale at the bottom shows time in MYA (Goodman 1996).

Independent fossil data suggest that the last common ancestor of humans within the chimpanzee-bonobo clade lived approximately 6-7 MYA (Brunet et al. 2002). The fossil record, however, consists predominantly of fossils of human ancestors and related species and has provided extensive data regarding the evolution of human ancestors (Wood and Collard 1999). In the case of the chimpanzee lineage, only one known example of a chimpanzee fossil record is known (McBrearty and Jablonski 2005).

Initial studies comparing humans and chimpanzees found very few genetic differences between the two species

Once it was established that the closest relatives of the humans were the chimpanzees, the genetic differences between the two species were explored. While the biochemical similarities between humans and chimpanzees have been noted for nearly a century, most of this evidence was based on immunological studies (Nutall 1904). In the second half of the twentieth century, comparative studies between humans and chimpanzees were done by electrophoretic analysis of blood proteins and by protein sequencing (Syner and Goodman 1966; Sarich and Wilson 1967; Doolittle et al. 1971; King and Wilson 1975). These studies helped us understand the genetic differences between the two species.

Initial studies comparing humans and chimpanzees sequences found very few differences between the two species

In a classic study (King and Wilson 1975) done in 1975, comparing the protein sequences of humans and chimpanzees, Mary-Claire King and Allan Wilson concluded that the humans and chimpanzees are 99% identical at the protein level and that the small degree of sequence divergence between the two species may not be sufficient to account for the phenotypic differences between the two species. They proposed that differences in gene regulation might be responsible for the phenotypic differences. Following this study, a number of follow-up studies (Doolittle et al. 1971; Eyre-Walker and Keightley 1999; Fujiyama et al. 2002) confirmed that the differences in amino-acid sequences between the two species is <1%.

Studies done by comparing direct samplings of chimpanzee genomic DNA with corresponding segments of the human genome (Britten 2002) indicate that 95% of the chimpanzee genome can be directly aligned with the human genome and that the sequence divergence in these precisely aligned regions is 1-2% (Fujiyama et al. 2002). The 5% divergence is due to insertions or deletions that have occurred in either genome since the divergence from the last common ancestor.

Other approaches found very few genetic differences between humans and chimpanzees

Other studies using different approaches have also reported very few genetic differences between humans and chimpanzees. The first genetic differences that became directly observable were the chromosomal differences between the two species, as highlighted by the karyotyping studies using a variety of dyes and fluorescent *in situ* hybridization (Yunis et al. 1980). The human-specific changes include a telomeric fusion of two ancestral chromosomes to form human chromosome 2 (chimpanzee chromosome 12 and 13) (Yunis et al. 1980); pericentric inversions of chromosome 1 and 18; and redistribution of heterochromatin on several chromosomes (Yunis et al. 1980; Yunis and Prakash 1982; Archidiacono et al. 1995; Meneveri et al. 1995).

An alternative to the King and Wilson gene regulation hypothesis (King and Wilson 1975) is the 'less-is-more-hypothesis', which emphasizes the loss of function of certain genes through gene inactivation events (Olson and Varki 2003). These inactivation events may have led to the evolution of a novel lineage such as the humans. A number of studies were done to identify gene inactivation events in both species. One such study reported

that unlike the chimpanzees, humans could not synthesize a form of cell surface sialic acid called *N*-glcolyl-neuraminic acid (Neu5Gc) (Chou et al. 1998). This sialic acid is widely expressed on the cell surfaces in many tissues in non-human primates. The synthesis of Neu5Gc depends on the hydroxylation of the precursor protein *N*-acetylneuraminic acid by CMP-Neu5Ac hydroxylase. In humans, the hydroxylase has been inactivated by a 92 bp deletion. This inactivation in humans could help explain the immunity that humans have to certain infections caused by viral and bacterial organisms that require Neu5Gc to recognize gut epithelial cells (Karlsson 1995; Varki 1997). Other known gene inactivation events in the human lineage involve the loss of one member of a large gene family, the members of which share overlapping functions. These gene inactivation events include the *V10* variable gene of the human T-cell-receptor-γ locus (Zhang et al. 1996; Rouquier et al. 1998), the olfactory receptor gene *OR 912-93* (Rouquier et al. 1998) and a type I hair-keratin gene (Winter et al. 2001).

Traditional ideas about genetic changes that lead to the evolution of new biological characteristics emphasize the role of gene duplication and divergence (Samonte and Eichler 2002). An example of a gene duplication event that is specific to humans is the gene that encodes protocadherin XY (Wu 2005). This gene lies in a region of the human Y chromosome that was duplicated and translocated from the X chromosome, after the last common ancestor of humans and chimpanzees. It has been postulated that protocadherin XY is involved in handedness, language lateralization and brain asymmetry (Crow 2002). This example of protocadherin XY shows that gene duplication

and divergence may be an important contributor to the phenotypic differences between the two species.

Other ideas of genetic differences between humans and chimpanzees include changes in the functional amino acids (Hellmann et al. 2003). An example of this that may have contributed to differences between humans and chimpanzees is seen in the highly conserved transcription factor forkhead box protein P2 (FOXP2*)* (Enard et al. 2002b; Zhang et al. 2002). Two amino-acid changes have been reported in FOXP2 in the human lineage, and genetic studies have implicated this protein in playing a role in human speech and language (Lai et al. 2001; Vargha-Khadem et al. 2005). The amino-acid changes in FOXP2 indicate that this protein may have been targeted for selection during recent human evolution.

Genome comparison studies confirm few genetic changes between humans and chimpanzees

Further clues to the differences between human and chimpanzees have come from whole genome comparison of human and chimpanzee genomes (Lander et al. 2001; Mikkelsen et al. 2005).

Initial observations from human whole genomic sequencing studies

With the advent of whole genome sequencing technologies, the initial draft of the human genome was published in 2001 (Lander et al. 2001; Venter et al. 2001). In-depth analyses of the chromosomes were done following the release of the initial draft, and in 2006, the

finished sequence of the human genome was made available (Muzny et al. 2006). Some of the major observations from this sequencing study were:

1. The human genome consists of 3.2 billion chemical nucleotides, i.e., adenine (Ad), guanine (Gu), thymine (Th) and cytosine (Cy).
2. The landscape of the human genome shows marked variation in the distribution of a number of features that include genes, transposable elements (mobile genetic elements that can move from one location to another location in the genome), CpG islands, GC content and recombination rates.
3. The number of genes in the human genome is 23,000-25,000.
4. Approximately 2% of the sequences in the human genome code for proteins.
5. Genes appear to be concentrated in random areas along the genome with vast expanses of non-coding DNA in between.
6. The number of single nucleotide polymorphisms (SNPs) identified is approximately 1.4 million.
7. Hundreds of genes appear to have resulted from horizontal transfer from bacteria at some point in the vertebrate lineage. Dozens of genes appear to have been derived from transposable elements.
8. Even though transposable elements make up approximately 50% of the human genome, there is a marked decline in the overall activity of these elements in the hominid lineage. DNA transposons and long-terminal repeat (LTR) retrotransposons may have become inactive in the human genome. Short interspersed nuclear elements (SINEs) make up 13% of the human genome and the Alu elements seem to be the most abundant SINE elements.

9. Segmental duplications are much more frequent in humans than in the yeast, fly or worm.

Human and chimpanzee whole genome comparisons show very few differences

Following the sequencing of the human genome, a logical follow-up was to sequence other related animals to better understand the evolutionary biology and unique features of humans. Given the near 99% similarity (King and Wilson 1975; Yunis et al. 1980; Britten 2002; Fujiyama et al. 2002) that humans share with the chimpanzees, there was a recommendation by the scientific community (McConkey and Varki 2000; Normile 2001; Olson and Varki 2003) to sequence the chimpanzee genome. By comparing the human and chimpanzee genomes, researchers felt that genes that contribute to human characteristics could be identified.

The initial draft of the chimpanzee genome was published in 2005 (Mikkelsen et al. 2005) and initial comparisons between the human and chimpanzee genomes have revealed the following observations:

1. About 30% of all human proteins are identical in sequence to their chimpanzee orthologs. Typical human and chimpanzee orthologs differ by only two amino acids on average.
2. The difference between human and chimpanzee genomes due to single nucleotide substitutions is 1.23%, of which 1.06% is due to fixed divergence and the rest is due to polymorphisms within either species.

3. Insertion and deletion (INDEL) events account for approximately 3% of the differences between the two species. The number of genetic changes due to INDELs, however, is fewer compared to the changes due to single nucleotide substitutions.
4. The rate of transposable element insertions is different between the two genomes, with the SINEs being threefold more active in the humans, whereas the chimpanzees have acquired two new families of retroviral elements.
5. Several loci in the human genome were identified as potential candidates for selective sweeps in recent human history.

Need for comparison with an out-group species

There is a problem in comparing the genomes of closely related species like the humans and chimpanzees. It is difficult to determine whether the genetic changes in either genome are due to changes in the human lineage or due to changes in the chimpanzee lineage, once the species diverged from their common ancestor. Comparing the human and chimpanzee genome sequences with a not-too-distantly related common ancestor (out-group species) would resolve this problem. Potential out-group genome species include the orangutan, which diverged from the human lineage 14 MYA, and the rhesus macaque, which diverged from the human lineage 25 MYA.

A draft sequence of the rhesus macaque (*Macaca mulatta)* – a type of Old world Monkey – was published in 2007 (Gibbs et al. 2007). Initial comparisons between human and rhesus macaque genomes have revealed that the sequence identity is approximately 93%, which is reduced to approximately 90%, if INDELs are taken into account. Like the

human and chimpanzee genomes, transposable elements constitute approximately 50% of the rhesus macaque genome, with the endogenous retroviruses having acquired some new families. The genomes of orangutan, gorilla and marmoset are being sequenced and are in early stages of completion (Varki 2007).

Human and chimpanzee gene expression comparison show an accelerated gene expression change in human brain evolution

As noted from independent (King and Wilson 1975; Yunis et al. 1980; Britten 2002; Fujiyama et al. 2002) and whole genome comparison studies (Mikkelsen et al. 2005), the sequence divergence between human and chimpanzee genomes is very little. It has been proposed that the phenotypic differences between the two species may be due to differences in gene regulation (King and Wilson 1975). Differences in gene regulation between humans and chimpanzees have been reported by microarray analysis studies (Enard et al. 2002a; Caceres et al. 2003; Khaitovich et al. 2004; Uddin et al. 2004). Most of these differences are reported in the brain.

Genes involved in control of brain size have different amino acid substitutions in the human lineage

One of the major anatomical differences between the human and chimpanzees is the exceptionally large size of the human brain (Figure 1.2). Estimates show that the human brain on average is three times as large as the chimpanzee brain (Carroll 2003; de Sousa and Wood 2007). The cerebral cortex, which is involved in memory, language and attention, is far larger in the humans than the chimpanzees and contains 50% more neurons (Preuss et al. 2004). It has been shown that two genes involved in the control of

brain size - abnormal spindle-like-microcephaly associated protein (ASPM) gene and the microcephalin gene - have an unusual pattern of amino acid substitution in the human lineage (Zhang 2003; Evans et al. 2004a; Evans et al. 2004b; Kouprina et al. 2004).

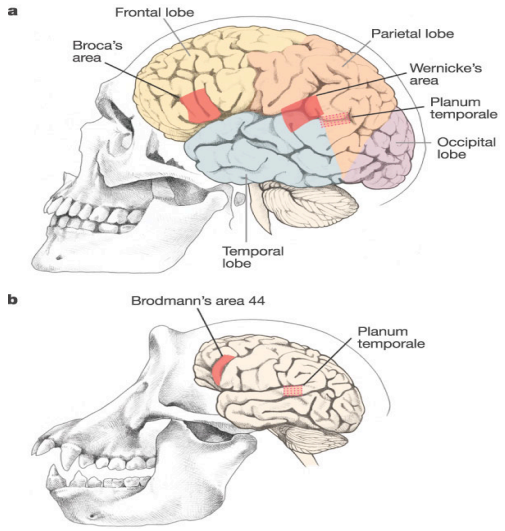


Figure 1.2: Comparison of the a) human and b) chimpanzee brains shows that humans have bigger brains with certain areas of specialized function (Carroll 2003).

Differences in brain function between humans and chimpanzees may be due to gene regulation differences

Differences in gene regulation between humans and chimpanzees can be studied by DNA microarray technology, which allows researchers to quantify the expression levels of thousands of genes simultaneously and to assess the role of gene-expression changes in evolution (Preuss et al. 2004). Much of our understanding of the differences in gene expression between human and chimpanzee brain function has come from microarray

studies (Enard et al. 2002a; Caceres et al. 2003; Karaman et al. 2003; Marvanova et al. 2003). Some of the major conclusions from these studies were:

1) The rate of gene expression changes in the brain accelerated during human evolution.
2) Gene expression changes in the evolution of the human brain primarily involved an up-regulation of expression.
3) Approximately 2-4% of the genes show differential expression between humans and chimpanzees in the cerebral cortex.

In one of the first studies done to understand the gene expression differences between human and chimpanzees (Enard et al. 2002a), the authors concluded that more pronounced changes in gene expression occurred in the human lineage, specifically in the brain. It was also shown that the gene expression changes in the liver accumulated at equal rates in both species, thus suggesting that the brain may be one of the tissues where expression changes are more profound. Later, in another study (Gu and Gu 2003), it was shown that the number of significant expression changes in the brain was threefold higher in humans than in chimpanzees.

From the above studies (Enard et al. 2002a; Caceres et al. 2003), it was also shown that majority of the genes that are differentially expressed in human and chimpanzee cerebral cortices show an increased level of expression in the human lineage. Conversely, gene expression analysis of liver and heart samples showed no evidence of an increased level of expression, with similar proportion of genes showing changes in expression in either

direction. A gene ontology analysis revealed that most of the genes showing an up-regulation in the human brain are involved in neuronal function and synaptic activity. However, Uddin and colleagues reported an increased down-regulation of genes in the human lineage, rather than an up-regulation (Uddin et al. 2004). Two of the flaws of this study (Uddin et al. 2004) were that the sample size of the individuals was small and that the down-regulation of expression could be explained by the fact that the focus of the results of this study was on small expression changes (Preuss et al. 2004).

In one of the most comprehensive microarray studies (Khaitovich et al. 2005) done to correlate the sequence differences between the two species with expression differences, it was shown that significant differences in gene expression patterns exist between humans and chimpanzees especially in organs (e.g., brain and testes) that can be correlated to difference in functions (e.g., cognitive ability and fertility).

One cautionary note about the microarray studies is the fact that the probes on the microarray are human-specific, and sequence mismatches between human probes and non-human primate mRNA can give a false indication that the gene is expressed at lower levels in the non-human primates than humans. In such cases, the data should be normalized, and the results should be confirmed by other techniques like Northern Blot or quantitative real time polymerase chain reaction (qrt-PCR). An alternate approach would be to discard those probe sets for which sequence mismatches are seen (Khaitovich et al. 2005).

Apoptosis is involved in both brain development and disease

It has been postulated that the most profound changes in human brain evolution have probably involved changes in expression of genes that function early in development (Preuss et al. 2004). Thus, a focus on the mechanisms of development may shed more light on the differences between human and chimpanzee brain evolution.

The mechanism of apoptosis

Programmed cell death (PCD) or apoptosis is considered as a vital component of various processes including normal cell turnover, proper development and functioning of the immune system and the development of the nervous system (Elmore 2007). Apoptosis is a tightly regulated mechanism and has many biochemical and cellular features associated with it.

One of the first cellular features of apoptosis that was identified was the fragmentation of DNA into nucleosomal fragments. This feature is used as a biochemical marker of apoptosis (Wyllie 1980). DNA fragmentation, however, is one of the end products of apoptosis, and a cell goes through various morphological changes during apoptosis (Figure 1.3).

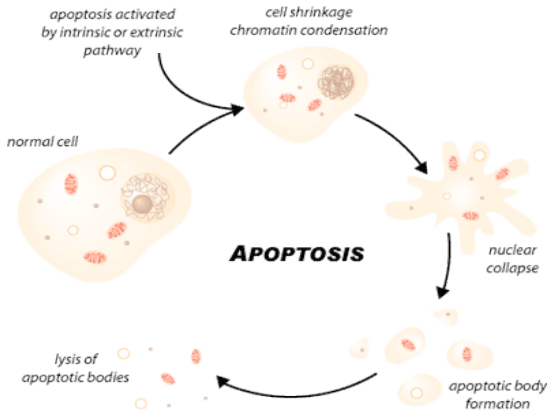


Figure 1.3: Morphological changes that occur in the cell during apoptosis (Elmore 2007).

The early stages of apoptosis are marked with cell shrinkage and chromatin condensation (pyknosis) (Kerr et al. 1972). This is followed by extensive plasma blebbing and nuclear membrane fragmentation (karyorrhexis). The cell fragments separate into apoptotic bodies during "budding". These apoptotic bodies are then phagocytosed by macrophages, parenchymal cells or neoplastic cells and degraded within the phagolysosomes. The cellular changes that occur during apoptosis can be identified by light and electron microscopy and are used to identify and distinguish apoptosis from necrosis. Necrosis is another form of cell death that involves an inflammatory reaction, followed by a release of cellular constituents into the surrounding interstitial tissue (Kerr et al. 1972; Elmore 2007).

Research indicates that there are two main apoptotic pathways: the extrinsic or death receptor pathway and the intrinsic or mitochondrial pathway (Kerr et al. 1972) (Figure 1.4). There is evidence that these two pathways are linked and that molecules of one pathway may influence molecules of the other pathway (Igney and Krammer 2002). Both

these pathways involve the activation of initiator and executioner caspases, and the pathways converge at the execution phase of the apoptotic cascade. The execution phase involves the activation of the executioner caspases and results in DNA fragmentation and phagocytosis.

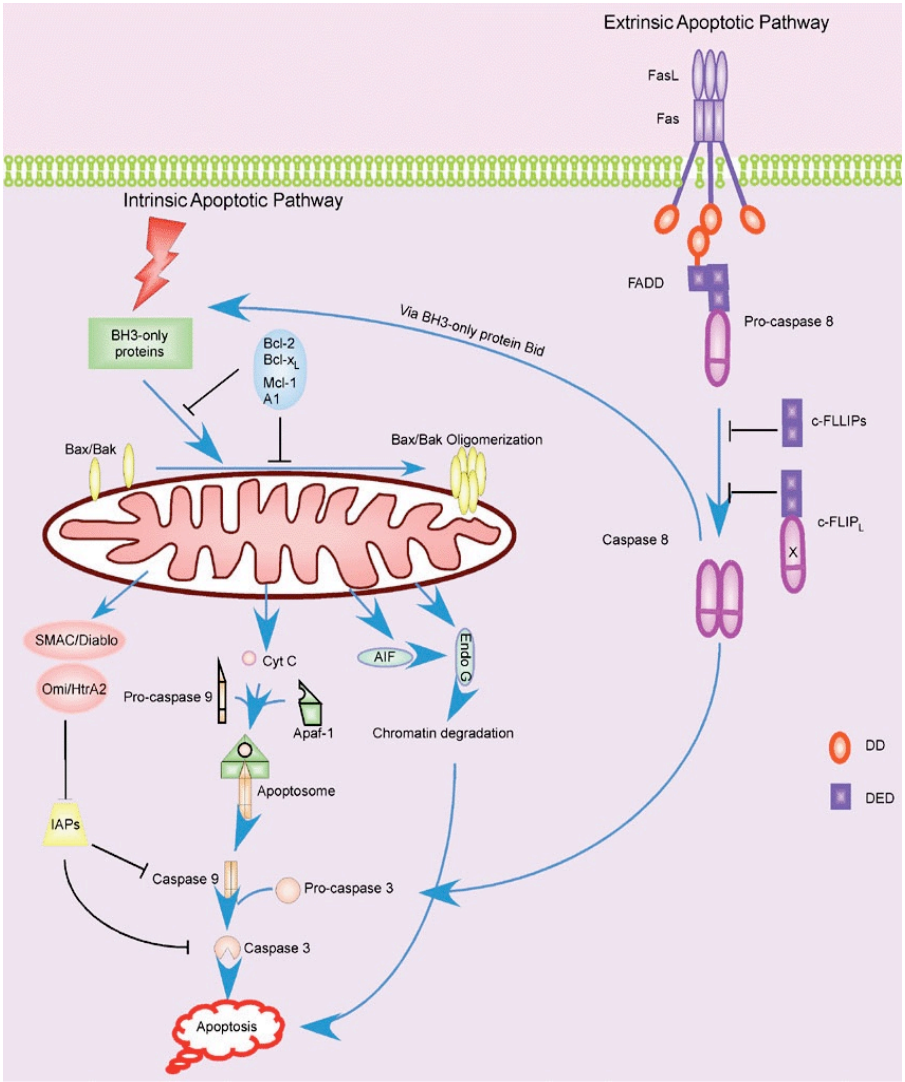


Figure 1.4: The two main apoptotic pathways-extrinsic and intrinsic (Putcha et al. 2002) (DD- death domain and DED-death effector domain)**.**

Caspases are a group of proteases, whose enzymatic properties are governed by a dominant specificity for substrates containing aspartate (Asp) residues. Caspases use a cysteine (Cys) side chain for their enzymatic activity (Salvesen and Dixit 1997; Thornberry et al. 1997). Caspases are further classified as initiator (caspase-8, -9, -10 and -2) and executioner caspases (caspase-3, -6 and -7), depending on their position and role in the apoptotic cascade (Salvesen 2002). In the cells caspases are present as inactive zymogens and are activated either by autocatalysis or by other proteins in the apoptotic cascade.

The extrinsic apoptotic pathway initiates apoptosis on the cell surface via death receptors. These receptors are members of the tumor necrosis factor (TNF) receptor gene super family (Locksley et al. 2001). Members of the TNF receptor super family have a cytoplasmic domain of 80 amino acids called the death domain (Ashkenazi and Dixit 1998), which is involved in transmission of the death signal from the cellular surface to the intracellular signaling pathways. The Fas ligand (FasL) and receptor (FasR) are one of the best-characterized ligand and receptors of the TNF super family (Chicheportiche et al. 1997). Binding of FasL with FasR results in the binding of Fas associated with death domain (FADD) protein with the ligand-receptor complex. FADD then associates with procaspase- 8, forming a death-inducing cell signaling complex (DISC), which results in the autocatalytic activation of procaspase-8 to caspase-8 (Kischkel et al. 1995). The activated initiator caspase-8 activates the execution phase of the apoptotic pathway. The extrinsic apoptotic pathway can be inhibited by cellular caspase-8 like inhibitory protein

(c-FLIP), which binds to FADD and caspase-8, rendering them ineffective (Kataoka et al. 1998; Scaffidi et al. 1999).

A number of intracellular signals such as stress or DNA damage initiate the intrinsic apoptotic pathway. These signals may act in a positive or negative fashion and act directly on targets within the cell, usually the mitochondria (Liu et al. 2003). The intracellular signals cause changes in the inner mitochondrial membrane that results in the opening of the mitochondrial permeability transition (MPT) pore and the release of a number of pro-apoptotic proteins, chief among them being the cytochrome C, second mitochondria-activator of caspases (Smac) or Diablo and the apoptosis inducing factor (AIF) proteins (Yang et al. 1997; Du et al. 2000; van Loo et al. 2002).

On release from the mitochondria, cytochrome C activates the apoptosis-activating factor 1 (Apaf1) as well as procaspase-9 to caspase-9, forming a multi-complex protein structure called the apoptosome (Chinnaiyan 1999; Hill et al. 2004). The apoptosome then activates the execution phase of the apoptotic pathway. The Smac/Diablo protein is a pro-apoptotic factor and prevents the inhibitor of apoptosis (IAP) group of proteins from inactivating the intrinsic pathway. On the other hand, AIF translocates to the nucleus and causes DNA fragmentation in a caspase independent manner (Joza et al. 2001).

The control and regulation of apoptotic mitochondrial events occur through members of the B-cell lymphoma 2 (Bcl-2) family of proteins (Cory and Adams 2002). Members of this family can either be pro-apoptotic, e.g., Bcl-2 associated X protein (Bax), BH3

interacting domain death agonist (Bid) and Bcl-2 associated death promoter (Bad), or they can be anti-apoptotic, e.g., Bcl-2, B-cell lymphoma-extra large (Bcl-x_L) and Bcl2-L2 (Bcl2-Like 2). It is thought that the Bcl-2 family of proteins alters the mitochondrial membrane permeability, which results in the release of cytochrome C from the mitochondria.

Both the extrinsic and intrinsic apoptotic pathways converge at the execution phase, which is considered as the final phase of the apoptotic pathway and involves the activation of the executioner caspases: caspase-3, -6 and -7. These caspases cleave various substrates, including cytokeratins and a DNA repair protein called poly ADP ribose polymerase (PARP), resulting in the biochemical changes seen in apoptotic cells (Slee et al. 2001).

Apoptosis in development

It has long been recognized that apoptosis plays a critical role in normal animal development (Elmore 2007). The developmental roles that apoptosis plays include the removal of unneeded structures and the shaping of existing structures to control cell number and organ size (Jacobson et al. 1997). The apoptotic process is highly conserved and orthologs of the various key apoptotic proteins are found in other organisms like yeast and fly (Meier et al. 2000).

Apoptosis has been known to occur during the normal development of the vertebrate nervous system (Oppenheim 1991). Traditional views of neural development have

focused on survival factors such as nerve growth factors and neurotrophins (Snider 1994). Traditionally, it was believed that during neural development, apoptosis is merely a mechanism to match neuronal populations to their target fields and that it plays a role in eliminating neurons with erroneous or inadequate projections (Cowan et al. 1984; Raff 1992). Recent studies have shown that apoptosis is an important mechanism involved in adjusting the initial pool of progenitor neuron cells for the proper morphogenesis of the nervous system (Cecconi et al. 1998; Yoshida et al. 1998; Kuan et al. 1999; Sabapathy et al. 1999).

It is known that in mammals, the intrinsic apoptotic pathway plays an important role in the formation of the nervous system (Haydar et al. 1999; Kuan et al. 2000; Putcha et al. 2002; Lossi and Merighi 2003). Most of the evidence for this has come from knockout studies done on mice. Mice carrying Apaf1 mutations display a failure of neural tube closure as well as an overgrowth of the forebrain (Cecconi et al. 1998; Yoshida et al. 1998; Honarpour et al. 2001). Caspase-9 knockout mice also show brain overgrowth, and this overgrowth appears as cortical folds (Kuida et al. 1998). Caspase-3 knockout mice also display neuronal overgrowth and disorganization (Kuida et al. 1996).

Studies extended to primates also show that the intrinsic apoptotic pathway plays a role in brain development, especially in the development of the neocortex (Rakic and Zecevic 2000; Chan et al. 2002). In a study done to understand the molecular evolution of the mammalian intrinsic apoptotic pathway, it was shown that Apaf1 and caspase-3 might have undergone accelerated evolution in the human lineage (Vallender and Lahn 2006).

Apoptosis in disease

Just like any other pathway, a misregulation of the apoptotic pathway can lead to a number of diseases, chief among them being neurodegenerative diseases (Nijhawan et al. 2000) and cancer (Lowe and Lin 2000). Some of the neurodegenerative diseases that exhibit a misregulation of the apoptotic pathway are Huntington's, Parkinson's and Alzheimer's. All of these diseases display a Mendelian form of inheritance and are associated with the loss of particular neuronal subtypes. In studies done on transgenic mice, it has been shown that in Huntington's disease, genes that code for caspase-1 and -3 are up-regulated during early as well as later stages of the disease (Ona et al. 1999; Chen et al. 2000). Studies have also shown that the up-regulation of the caspases is followed by a release of cytochrome C (Ona et al. 1999). An up-regulation of the caspases and a related increased apoptotic turnover has also been shown in Parkinson's disease (Klivenyi et al. 1999). In summary, an overall up-regulation of key apoptotic genes is seen in neurodegenerative diseases.

In cancer, the apoptotic machinery is disrupted, resulting in tumor initiation, progression and metastasis (Lowe and Lin 2000). The tumor suppressor protein p53 is known to up-regulate the expression of pro-apoptotic gene Bax in response to DNA damage and other stress signals. Bax in turn stimulates the mitochondria to release cytochrome C, and this triggers apoptosis. Consequently, a number of sub cellular structures and organelles are destroyed, including the ones where there is any form of stress or DNA damage. Thus, the fidelity of the genome is preserved. In a number of tumors, however, it has been

shown that p53 is mutated and its tumor suppressor function is lost (Wallace-Brodeur and Lowe 1999). This may explain the evasion of apoptosis in certain cancers.

It has been shown by transgenic mice studies that the anti-apoptotic genes Bcl-2 and Bcl-x_L may be over-expressed in certain tumors (Vaux et al. 1988; McDonnell et al. 1989; Hockenbery et al. 1990). Bcl-2 promotes survival of cells through its anti-apoptotic properties. Conversely, it has been shown pro-apoptotic Bax is inactivated in certain colon cancers (Rampino et al. 1997; Meijerink et al. 1998).

Humans show a higher susceptibility to cancer than chimpanzees (Seibold 1973; Puente et al. 2006). In humans, the deaths caused by certain neoplasms like breast, prostate and lung are more than 20%, compared to the incidences of these cancers in chimpanzees which is ~ 2% (McClure 1973; Seibold 1973; Beniashvili 1989). It was reported in a study that the chimpanzee gene encoding for p53 contains a proline (Pro) at codon 72, while in the humans this codon is polymorphic and can code for arginine (Arg) and Pro (Puente et al. 2006). In a comparison of the p53 sequence in a number of different primates, it was seen that the codon 72 codes for Pro, thus suggesting that this must be the ancestral allele and that the Arg72 allele must be unique to the human lineage. In the same study, it was shown that the breast cancer 1 (BRCA1) gene in the chimpanzee has a 8kb deletion, the consequences of which are not known.

From the above discussion, it is seen that the apoptotic pathway is involved in the development of the brain. Further, it has been shown that the human cerebral cortex

contains 50% more neurons than the chimpanzees (Preuss et al. 2004). In a study done on mice, it was shown that by disrupting the function of the caspases-3 and -9 in the brain, a decrease in the apoptotic function was observed resulting in an expansion of the forebrain (Kuan et al. 2000). The expansion of the forebrain was due to the survival of a specific type of forebrain progenitor cells.

One question that arises is whether the difference in the number of neurons between humans and chimpanzees can be attributed to a difference in the apoptotic function between the two species as caused by one of the genes in the apoptotic pathway. If such a difference does exist, can it also explain the difference in susceptibilities that humans and chimpanzees have to diseases like cancer? Not much is known about the differences in the apoptotic function between the two species, and answers to these questions may help us better understand the differences in brain development and function as well as the difference in the susceptibilities that humans and chimpanzees have to diseases associated with apoptosis.

Impact of INDELs on human and chimpanzee genomes

From independent (Britten 2002) and sequencing studies (Mikkelsen et al. 2005), it was shown that INDELs account for approximately 4% of the differences between human and chimpanzee genomes and that INDELs may contribute to the phenotypic differences between the two species. INDELs consist of a number of nucleotides that are present in the genome (INsertion) of one organism but may be absent from the genome of a closely related species (DELetion). By using an out-group species, INDELs can be characterized more precisely.

Previous studies done to identify INDELs

In one of the first studies done to identify human and chimpanzee INDEL variation, a 27 Mb region of the human chromosome 21 was compared with chimpanzee DNA sequences, and approximately 57 INDELs were identified (Frazer et al. 2003). In the same study, a 9 Mb region of the human chromosome 21 was compared with DNA sequences from orangutan, rhesus macaque and wholly monkey, and 114 INDELs were identified.

In a comparison of the high quality sequence of the chimpanzee chromosome 21 and the human chromosome 22, as many as 68,000 INDELs were identified (Watanabe et al. 2004). Most of these INDELs were associated with SINE elements. Later on, in another study, small INDELs of size ≤ 100 bp between the human chromosome 21 and the chimpanzee chromosome 22 were identified (Chen et al. 2007). This study used human-chimpanzee-mouse-dog multiple sequence alignment to identify the human-specific INDELs. Most of the INDELs identified in this study are located in the intergenic region of the genome.

In a recent comparison of human chromosome 21 and chimpanzee chromosome 22, as many as 6279 INDELs of size >10 bp were identified (Volfovsky et al. 2009). Many of these INDELs are located in the intron region of the genes. From the above studies, it has been shown that INDELs are made of transposable elements (TEs) and tandem repeats (Mills et al. 2006a).

Transposable elements and their impact on human chimpanzee genome evolution

From sequencing (Lander et al. 2001; Venter et al. 2001; Mikkelsen et al. 2005; Gibbs et al. 2007) as well as independent studies (Bonner et al. 1982; Holmes et al. 1994), it was shown that approximately 50% of the primate genomes contain TEs. It has also been recognized that TEs may be important factors in regulatory evolution and hence have the capacity to alter gene expression (McClintock 1984; Bowen and Jordan 2002; van de Lagemaat et al. 2003) among other functions.

First discovered in 1944 by Barbara McClintock (Mc 1950), TEs were initially thought to be genomic parasites. It was observed that these elements could out-replicate and spread in natural populations with a selective disadvantage to the host organisms (Hickey 1982). However, later on, an adaptive role of these elements in the host genomes emerged (McDonald 1993; Kidwell and Lisch 2000; Kidwell and Lisch 2001). It was shown that TEs are important contributors in *Drosophila* telomere maintenance (Pardue et al. 1987) and in mammalian DNA repair (Morrish et al. 2002). A new role of these elements in the evolution of epigenetic mechanisms is now emerging (Finnegan 1992; Huda and Jordan 2009; Huda et al. 2009).

TEs can be classified on the basis of their mechanism of movement in the host genomes: as Class I elements that consists of retrotransposons, or as Class II elements that consists of DNA transposons (Finnegan 1992). Retrotransposons move throughout the genome via an RNA intermediate, whereas the DNA transposons move through a “cut and paste” mechanism (Figure 1.5).

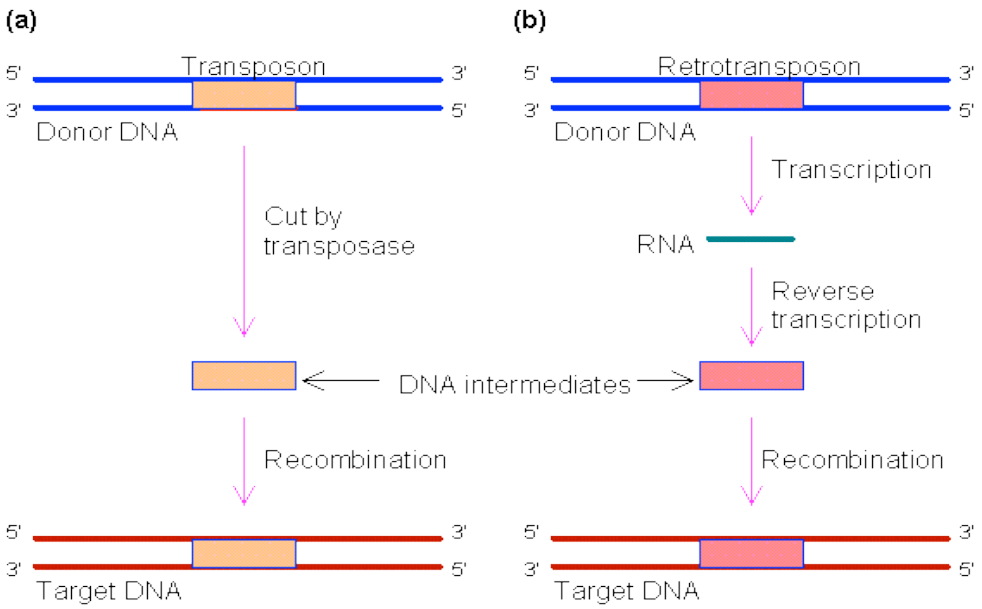


Figure 1.5: Mechanism of movement of a) DNA transposon and b) Retrotransposon (Deininger and Batzer 2002).

Class I elements are further classified into two groups: a) LTR retrotransposons/ endogenous retroviruses and b) non-LTR retrotransposons (Figure1.6) (Deininger and Batzer 2002). The non-LTR retrotransposons consist of long interspersed nuclear elements (LINEs) and SINEs. The LTR retrotransposon/ endogenous retroviruses and LINEs are autonomous in the sense that they encode their own reverse transcriptase enzyme for transcription, whereas the SINEs depend on the reverse transcriptase from the LINEs and hence are non- autonomous. For a long time, it has been known that retrotransposons have played an essential role in the evolution of mammalian gene expression (Sverdlov 2000; Medstrand et al. 2005).

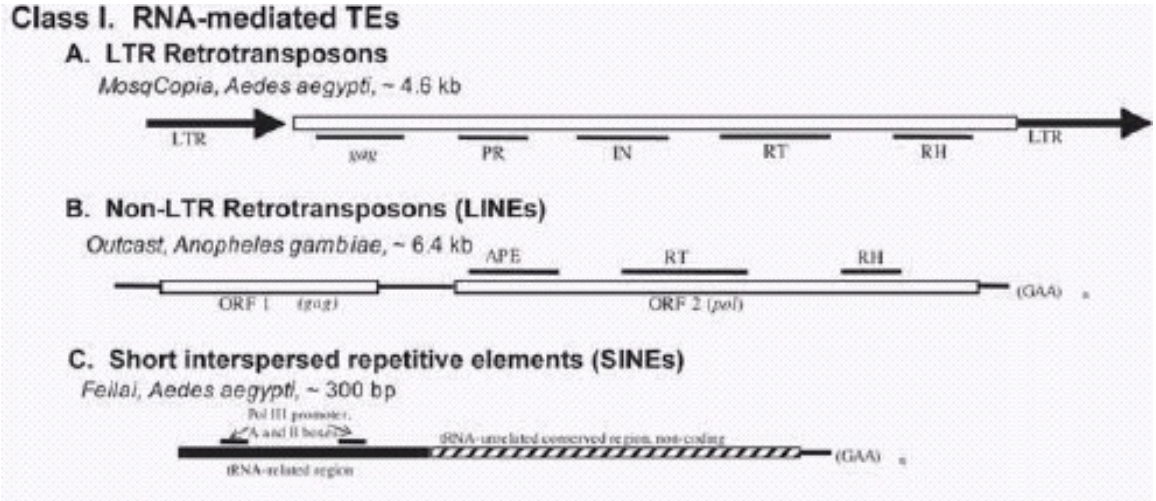


Figure 1.6: Classification of Retrotransposons (Kapitonov and Jurka 2003).

From sequencing studies, it is known that a particular class of SINE elements- the Alu element- is threefold more active in the human genome than in the chimpanzee genome, whereas the chimpanzees have acquired two new families of endogenous retroviruses (Mikkelsen et al. 2005). In a study done to classify recently inserted transposable elements between humans and chimpanzees, it was shown that humans have a greater rate of insertions compared to chimpanzees, and the majority of these insertions are in or around genes (Mills et al. 2006a). Most of these insertions were associated with Alu elements. Further, in a previous study from our lab (Polavarapu et al. 2006), it was shown that transposable elements, in particular retrotransposons, have contributed significantly to the INDEL variation that exists between humans and chimpanzees.

Tandem repeats and INDEL variation between humans and chimpanzees

Tandem repeats are adjacent patterns of two or more nucleotides that are repeated throughout the DNA (Benson 1999) and consist ~3% of the human genome (Lander et al. 2001).

In recent years the role of trinucleotide repeats in diseases like Huntington's (Trottier et al. 1994), fragile-X mental retardation (Verkerk et al. 1991) and myotonic dystrophy (Fu et al. 1992) has increased the interest in tandem repeats. It is known that tandem repeats also play roles in the development of the immune system cells (Verstrepen et al. 2004), and their potential roles in the evolution of social-sexual behavior among the humans and bonobos are now emerging (Hammock and Young 2005).

Tandem repeats have been known to play roles in the regulation of gene expression by either interacting with transcription factors, altering chromatin structure or acting as potential protein binding sites (Hamada et al. 1984; Pardue et al. 1987; Yee et al. 1991). They are known to contribute to the INDEL variation that exists between humans and chimpanzees (Madsen et al. 2008).

From the above discussion, it is clear that INDELs contribute to the variation between human and chimpanzee genomes. INDELs made up of either TEs or tandem repeats are usually located in or near genes. The role that INDELs play in differential gene expression between humans and chimpanzees has not been studied in detail. Given the very little sequence divergence between the two species, correlating INDEL variation

with differential gene expression may help us better understand the genetic basis of the phenotypic differences between the two species.

Analysis of the differences in apoptotic function between humans and chimpanzees and a role that INDELs may play in gene regulation differences between the two species.

One of the major phenotypic difference between humans and chimpanzees is the exceptionally large human brains and the associated higher cognitive functions (Gagneux and Varki 2001; Hacia 2001; Carroll 2003; de Sousa and Wood 2007; Varki 2007). In addition humans have a high susceptibilty to neurodegenerative diseases like Huntington's and Alzheimer's (Gearing et al. 1994; Poduri et al. 1994; Walker 1999) and other diseases like cancer (Lowe and Lin 2000). One pathway that is common to both the development of the brain and these diseases is the apoptotic pathway (Kerr et al. 1972; Elmore 2007). It has been proposed that an understanding of the mechanisms involved in the development of the brain may shed more light on the human chimpanzee brain evolution (Preuss et al. 2004). This dissertation looks at the differences in the apoptotic function between humans and chimpanzees and uses both computational and experimental tools to assess these differences.

Chapter 2 details a re-analysis of a previous microarray study (Khaitovich et al. 2005). The microarray expression data for this previous study was obtained from five chimpanzee and six human tissue samples. In the re-analysis the known expression differences between humans and chimpanzees are overlaid on pathways like apoptosis to

see whether a difference in apoptotic function is seen between the two species. The results from this analysis suggest that the apoptotic function may be reduced in humans compared to chimpanzees.

In chapter 3, the hypothesis that the apoptotic function may be reduced in humans is tested on human, chimpanzee and macaque primary fibroblasts cells, using a number of apoptotic assays and techniques. The results are consistent with the differential apoptotic hypothesis that humans have a reduced apoptotic function, compared to both chimpanzees and macaques.

Chapter 4 describes a study where the INDEL variation between humans and chimpanzees is identified. This is the first study that uses macaque genome sequence as an out-group to characterize the INDEL variation between humans and chimpanzees and correlates the INDEL variation to differences in gene regulation between the two species.

CHAPTER 2

DID NATURAL SELECTION FOR INCREASED COGNITIVE ABILITY IN HUMANS LEAD TO AN ELEVATED RISK OF CANCER?

SUMMARY

Despite the overall genetic similarity that exists between humans and chimpanzees, the species are phenotypically distinct. Among the most notable distinctions are differences in brain size and cognitive abilities. Previous studies have shown that significant differences in gene expression exist between the human and chimpanzee brain. Integration of currently available gene expression data with known metabolic and signaling pathways indicates that the expression of genes involved in the programmed cell death of brain neurons is significantly different between humans and chimpanzees and predictive of a reduced level of neuron apoptosis in the human brain. This pattern of expression is generally maintained in other human organs suggesting that apoptosis is reduced in humans relative to chimpanzees. We propose that a decreased rate of programmed neuron death may have been a consequence of selection for increased cognitive ability in humans. Since reduced apoptotic function is associated with an increased risk of cancer and related diseases, we hypothesize that selection for increased cognitive ability in humans coincidently resulted in an increased risk of cancer and other diseases associated with reduced apoptotic function.

INTRODUCTION

Over the ~6 million years that the human and chimpanzee lineages have diverged from a common ancestor, the two species have evolved a variety of distinctive phenotypic traits (Graves 1986; Wood and Collard 1999). Some of these differences are obviously adaptive, e.g., the larger size of the human brain and associated increase in cognitive abilities (Waters et al. 1998; Preuss 2000; Williams 2002; Herrmann et al. 2007) while others are less easily explained, e.g., the fact that humans are inherently more prone to developing cancer than chimpanzees (McClure 1973; Seibold 1973; Beniashvili 1989; Waters et al. 1998).

In an effort to understand the molecular bases of the phenotypic differences that distinguish humans from chimpanzees, a number of comparative genomic studies have been conducted in recent years (Enard et al. 2002a; Yan et al. 2002; Preuss et al. 2004; Khaitovich et al. 2005). For example, Khaitovich et al. (Khaitovich et al. 2005) have shown that significant differences in gene expression patterns exist between humans and chimpanzees in tissues (e.g., brain and testes) associated with traits (e.g., cognitive ability and fertility) distinguishing the two species. Our pathway reanalysis of these data suggests that genes involved in apoptotic function are generally lower in expression in humans relative to chimpanzees. In this paper, we discuss the possible significance of these findings and hypothesize how they may help explain the relative increased risk of cancer in humans.

HYPOTHESIS

Comparisons of gene expression patterns between human and chimpanzee brains suggest that increased cognitive ability in humans may be explained, at least in part, by a reduction in the level of programmed cell death in the human brain. Similar differences in gene expression patterns in other tissues indicate that the relative reduction in apoptotic function in humans may extend beyond the brain to other tissues. It has been previously noted that cancer and other diseases associated with reduced apoptotic function rarely occur in chimpanzees (McClure 1973; Seibold 1973; Beniashvili 1989; Waters et al. 1998). We hypothesize that these two observations may be linked and that a bi-product of natural selection for increased brain size and cognitive ability in humans was an elevated propensity for cancer and other diseases associated with reduced apoptotic function.

EVALUATION OF THE HYPOTHESIS

Empirical data

An extensive analysis of gene expression patterns between humans and chimpanzees was conducted by Khaitovich et al.(Khaitovich et al. 2005). A major goal of this previous study was to correlate sequence differences with expression differences and a number of microarray probe sets for which quality sequence could not be obtained in humans and chimpanzees (e.g., required for the calculation of ka/ks values) were excluded. Since the quality of the chimpanzee genome sequence has improved in recent years, and because our focus of interest is on the potential adaptive significance of chimpanzee–human expression differences, we reanalyzed this microarray dataset and incorporated probe sets

previously excluded.

The most dramatic difference in gene expression between humans and chimpanzees is in testis (62% of genes display a significant difference in expression) followed by heart (35%), brain (34%), kidney (33%) and liver (25%) (Table 2.1). To better understand the possible significance of the gene expression differences in the brains of the two species, we integrated the microarray data with known metabolic and signaling pathways. The results indicate that a number of the genes differentially expressed between the human and chimpanzee brains have been previously associated with human cognitive diseases including Parkinson's disease and Huntington's Chorea (TableA.1, Appendix A). It has been previously shown that the expression of genes predicted to induce apoptosis are elevated in the brains of cognitively impaired Parkinson's, Huntington's and Alzheimer's patients and in experimental (mouse) models of these diseases (Nijhawan et al. 2000; Cadet et al. 2001; Jayanthi et al. 2001; Kiechle et al. 2002; Reddy et al. 2004; Manczak et al. 2005).

Table 2.1:Differentially expressed genes between humans and chimpanzees across five organs. The human–chimpanzee gene expression data (.cel files) for five different tissues (i.e. brain, heart, liver, kidney and testis) in six humans and five chimpanzees were kindly provided (Khaitovich et al. 2005). The data were processed using the MAS normalization method encoded in the Affymetrix function library of the Bioconductor package (http://www.bioconductor.org/) developed for R statistical programming environment (http://www.r-project.org/). To insure the validity of the comparative gene expression levels, probes with sequences that differ between humans and chimpanzees were excluded from the analysis as previously described (Khaitovich et al. 2005). Only genes with detection p-values of less than 0.065 were considered for further analysis. Gene expression values were normalized across samples by Z-score calculation using Spotfire Decision Site software (http://spotfire.tibco.com/index.cfm). Genes with t-test, p-values of less than 0.01 between human and chimpanzees were considered differentially expressed.

	Brain	**Heart**	**Liver**	**Kidney**	**Testis**
Detected	10231	9580	9451	10546	12509
Dif Exp	3519	3361	2385	3458	7783
Percentage	34.39	35.08	25.23	32.78	62.21

To directly explore the possible consequence of the differences in gene expression between humans and chimpanzees on apoptotic function, we overlaid the gene expression differences between chimpanzee and human brains onto the human apoptotic-signaling pathway. Thirty-seven of the genes differentially expressed between human and chimpanzee brains were found to be components of the extrinsic or intrinsic apoptotic pathway (Tables 2.2 and 2.3, Figure 2.1). For example, cytochrome C (CYCS), a gene that is significantly down-regulated in the human brain, is the major initiator in the caspase-dependent intrinsic apoptotic pathway. A gene set enrichment analysis (GSEA) (http://www.broad.mit.edu/gsea/) further confirmed that our set of differentially expressed genes is significantly enriched for genes previously associated with apoptosis (p-value = $1.21 * 10^{-16}$).

Table 2.2: Genes differentially expressed between human and chimpanzee brains previously associated with the apoptotic pathway. The postulated function of each encoded protein in the apoptotic pathway (I = inducer of apoptosis; R = repressor of apoptosis) was obtained from the Ingenuity gene explorer annotations (S) (http://www.ingenuity.com/) and independently verified using the annotations at The Atlas of Genetics and Cytogenetics in Oncology and Haematology (A) (http://atlasgeneticsoncology.org/index.html) and/or by literature search.

Gene Symbol (alias)	Gene Name	Location in Human Genome	Function
1.AIFM1	Apoptosis inducing factor mitochondrion associated, 1	chrXq25-q26	I[S, A]
2.BID	BH3 interacting domain death agonist	chr22q11.1	I[S, A]
3.CAPN2	Calpain 2	chr1q41-q42	I[S, A]
4.CAPN3	Calpain 3	chr15q15.1-q21.1	I[S, A]
5.CAPN5	Calpain 5	chr11q14	I[S, A]
6.CAPN7	Calpain 7	chr3p24	I[S, A]
7.CASP2	Caspase 2, apoptosis related cysteine protease	chr7q34-q35	I[S, A]
8.CASP6	Caspase 6, apoptosis related cysteine protease	chr4q25	I[S, A]
9.CASP9	Caspase 9, apoptosis related cysteine protease	chr1q36.3-p36.1	I[S, A]
10.CYCS	Cytochrome C, somatic	chr7p15.3	I[S, A]
11.DIABLO	Diablo homolog (Drosophila)	chr12q24.31	I[S, A]
12.HTRA2	HtrA serine peptidase 2	chr2p12	I[S, A]
13. MKK7 (MAP2K7)	Mitogen-activated protein kinase kinase 2	chr19p13.3-p13.2	I[S, A]
14.ASK1 (MAP3K5)	Mitogen –activated protein kinase kinase kinase 3	chr6q22.33	I[S, A]
15.MAP4K4	Mitogen-activated protein kinase kinase kinase kinase 4	chr2q11.2-q12	I[33]
16.ERK2 (MAPK1)	Mitogen-activated protein kinase 1	chr22q11.2	I[34]

Table 2.2 continued

17.ERK1 (MAPK3)	Mitogen-activated protein kinase 3	chr16p11.2	I^{34}
18. JNK (MAPK8)	Mitogen-activated protein kinase 8	chr10q11.2	$I^{S, A}$
19.R-RAS (MRAS)	Muscle RAS oncogene homolog	chr3q22.3	I^{35}
			$I^{S,A}$
20.PRKCE	Protein kinase C,epsilon	chr2p21	$I^{S,A}$
21.SPTAN1	Spectrin alpha, non-erythrocytic 1 (alpha fodrin)	chr9q33-q34	$R^{S,A}$
22.XIAP (BIRC4)	Baculoviral IAP repeat-containing 4	chrXq25	$R^{S,A}$
23.IKKA (CHUK)	Conserved helix-loop-helix ubiquitous kinase	chr10q24-q25	$R^{S,A}$
24.DDR1	Discoidin domain receptor tyrosine kinase 1	chr6p21.3	$R^{S,A}$
25.DDR2	Discoidin domain receptor tyrosine kinase 2	chr1q23.3	$R^{S,A}$
26. ICAD (DFFA)	DNA fragmentation factor, 45 kDa ,alpha polypeptide	chr1q36.3-p36.2	$R^{S,A}$
27. KGFR (FGFR2)	Fibroblast growth factor receptor 2	chr10q26	$R^{S,A}$
28.KRAS	v-ki-ras2 Kirsten rat sarcoma viral oncogene homolog	chr12p12.1	$R^{S, A}$
29. MEK2 (MAP2K2)	Mitogen activated protein kinase kinase 2	chr19p13.1	$R^{S,A}$
30.RAF1	v-raf-1 murine leukemia viral oncogene homolog 1	chr3p24	$R^{S, A}$
31. NFκB (RELA)	Nuclear factor of kappa light polypeptide gene enhancer in B-cells 3	chr11q-q13	$R^{S, A}$
32. TEK	TEK tyrosine kinase, endothelial	chr9p21	R^{36}
33.EGFR	Epidermal growth factor receptor	chr7p12	R^{37}

Table 2.2 continued

34.IKBKB	Inhibitor of kappa light polypeptide gene enhancer in B cells, kinase B	chr8p11.2	$R^{S, A}$
35.PARP1	Poly (ADP-ribose) polymerase family member 1	chr1q41-q42	$R^{S, A}$
36.TYRO3	TYRO3 protein tyrosine kinase	chr15q15.1-q15.2	$R^{S,A}$
37.PDGFRA	Platelet-derived growth factor receptor, alpha polypeptide	chr4q11-q13	$R^{S,A}$

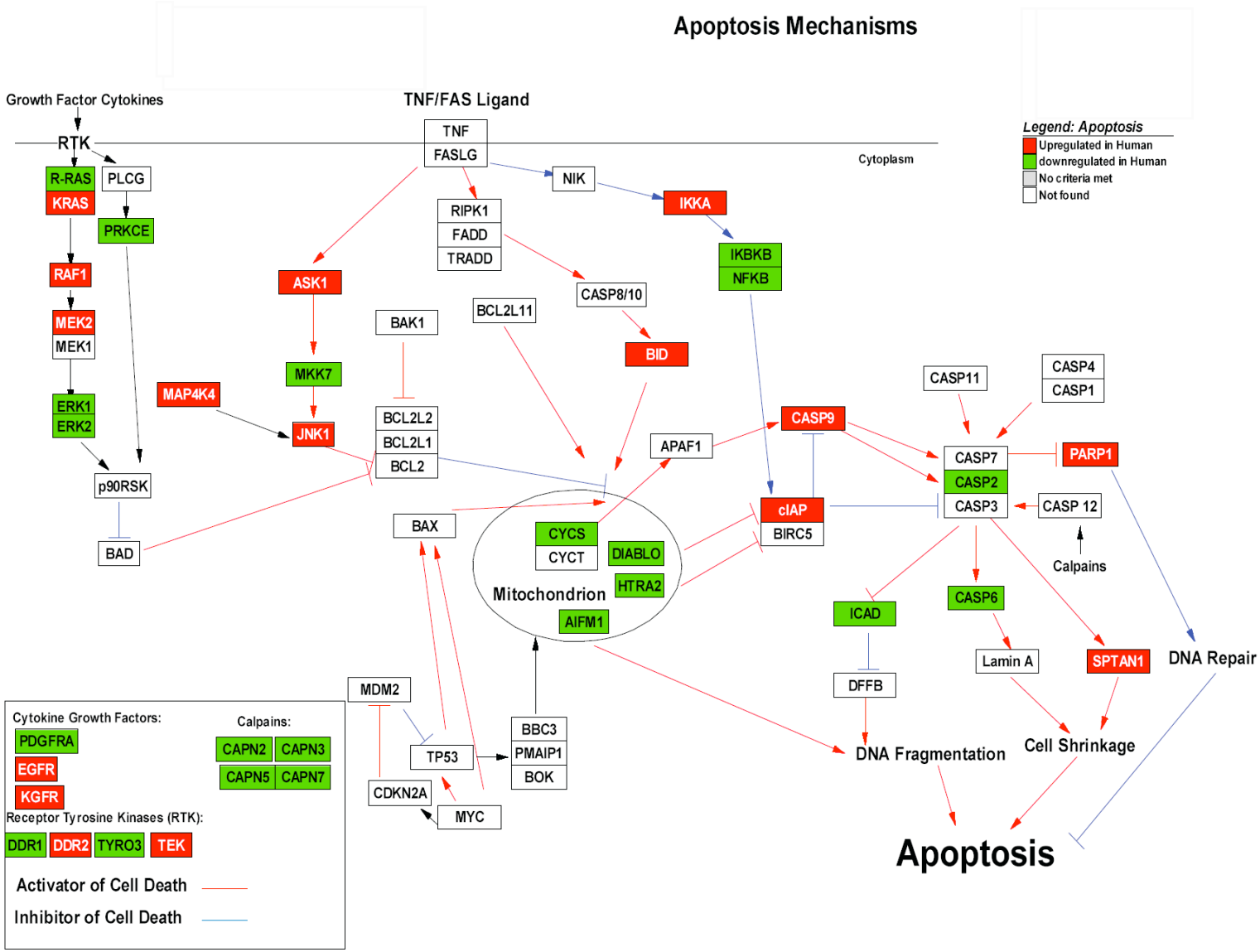

Apoptosis Mechanisms
Growth Factor Cytokines
TNF/FAS Ligand
RTK
TNF
FASLG
Cytoplasm
Legend: Apoptosis
Upregulated in Human
downregulated in Human
No criteria met
Not found
R-RAS
KRAS
PLCG
PRKCE
RAF1
MEK2
MEK1
ERK1
ERK2
p90RSK
BAD
ASK1
MKK7
MAP4K4
JNK1
RIPK1
FADD
TRADD
NIK
IKKA
IKBKB
NFKB
CASP8/10
BID
BAK1
BCL2L11
BCL2L2
BCL2L1
BCL2
BAX
CYCS
CYCT
DIABLO
HTRA2
AIFM1
Mitochondrion
APAF1
CASP9
cIAP
BIRC5
CASP11
CASP4
CASP1
CASP7
CASP2
CASP3
PARP1
CASP 12
Calpains
CASP6
ICAD
DFFB
Lamin A
SPTAN1
DNA Repair
DNA Fragmentation
Cell Shrinkage
Apoptosis
MDM2
TP53
CDKN2A
MYC
BBC3
PMAIP1
BOK
Cytokine Growth Factors:
PDGFRA
EGFR
KGFR
Receptor Tyrosine Kinases (RTK):
DDR1
DDR2
TYRO3
TEK
Calpains:
CAPN2
CAPN3
CAPN5
CAPN7
Activator of Cell Death
Inhibitor of Cell Death

Figure 2.1: Apoptotic pathway genes are differentially expressed in the human and chimpanzee brains consistent with a model of reduced neuron cell death in the human brain. Red nodes indicate genes displaying relatively higher levels of expression ($p < 0.01$) in human brain; green nodes indicate genes displaying relatively lower levels of expression ($p < 0.01$) in human brain. Pathways were constructed using Ingenuity®Systems (http://www.ingenuity.com/) and modified using GenMAPP (http://www.genmapp.org).

Twenty-one of the 37 apoptotic genes displaying differences in expression between human and chimpanzee brains have been characterized previously as inducers of apoptosis and 16 (76%) of these are down-regulated in the human brain (Table 2.3). Sixteen of the 37 genes have been characterized previously as repressors of apoptosis and 9 (56%) of these genes are up-regulated in the human brain. It has been proposed that natural selection for increased cognitive ability in humans resulted in an elevated rate of brain neuron synthesis (Williams 2002). Our findings are consistent with the hypothesis that a decreased rate of programmed neuron death may have been an additional consequence of selection for increased cognitive ability in humans.

Table 2.3: Differences in the expression of apoptotic pathway genes between human and chimpanzee brains are conserved across tissues. Shown are log-ratio values (average human log expression value minus average chimpanzee log expression value) of gene expression differences in human relative to chimpanzee. A negative value indicates down-regulation in humans relative to chimpanzees, while a positive value indicates up-regulation in humans relative to chimpanzee. The function of each encoded protein in the apoptotic pathway (I = inducer of apoptosis; R = repressor of apoptosis) was obtained from Ingenuity gene explorer annotations (Ingenuity® Systems, http://www.ingenuity.com/) and independently verified using the annotations at The Atlas of Genetics and Cytogenetics in Oncology and Haematology (http://atlasgeneticsoncology.org/index.html) and/or by literature search (see Table 2.2 for details). The direction of brain expression (+/-) differences with differences in expression in other tissues was highly significant by Signs test ($p < 6.15 * 10^{-11}$).

Gene Symbol	Function	Heart	Kidney	Liver	Testis	Brain
1.AIFM1	I	-0.07	-0.35	-0.0056	0.48	-0.79
2.BID	I	0.25	-0.48	0.21	0.86	0.56
3.CAPN2	I	-0.63	-0.95	-0.56	-0.29	-0.63
4.CAPN3	I	-1.68	1.45	1.4	-0.67	1.22
5.CAPN5	I	-0.09	-0.59	0.74	0.89	-0.51
6.CAPN7	I	-0.37	-1.06	-0.39	-0.28	-0.39
7.CASP2	I	-1.52	-1.05	-0.16	-0.48	-1.44
8.CASP6	I	-1.38	-1.33	-1.53	-0.48	-1.78
9.CASP9	I	0.14	0.23	-0.12	-0.031	-0.35
10.CYCS	I	-0.52	-0.95	-0.7	-0.33	-1.18
11.DIABLO	I	-0.79	-0.36	-0.12	-0.68	-0.51
12.HTRA2	I	-0.19	-0.43	-0.14	-0.73	-0.39
13.MKK7	I	-0.53	0.02	-0.25	0.26	-0.47
14.ASK1	I	0.81	-0.33	0.53	0.78	0.75
15.MAP4K4	I	0.33	0.21	0.77	1.64	0.89
16.ERK1	I	-0.39	-0.44	-0.17	0.036	-0.57
17.ERK2	I	-0.5	-0.05	-0.53	1.3	0.68
18.JNK1	I	0.02	0.52	0.82	0.92	-0.51
19.R-RAS	I	-0.27	-1.66	-0.88	-0.42	-1.14
20.PRKCE	I	-0.74	-1.59	-0.11	-1.38	-0.74
21.SPTAN1	I	0.11	0.12	-0.65	0.69	-0.14
22.IAP	R	1.75	1.02	0.53	1.87	2.08
23.IKKA	R	0.23	0.18	0.47	0.67	0.53
24.DDR1	R	-1.09	0.35	0.62	0.87	0.5
25.DDR2	R	0.35	0.4	0.09	0.76	0.56
26.ICAD	R	-0.41	-0.39	-0.12	0.09	-0.62
27.FGFR2	R	0.23	-0.58	0.235	0.051	1.38
28.KRAS	R	0.035	0.35	0.61	0.77	0.93
29.MEK2	R	1.17	0.65	-0.66	1.24	1.14
30.RAF1	R	0.21	-0.1	-0.61	0.07	-0.46
31.NFκB	R	-0.66	0.039	-0.23	0.29	-0.63
32. TEK	R	0.65	0.58	0.68	0.6	0.41
33.EGFR	R	0.07	1.02	0.75	1.81	0.87
34.IKBKB	R	-0.42	-0.33	-1.77	0.33	-0.52
35.PARP1	R	0.94	-0.33	-0.212	0.3	-0.32
36.TYRO3	R	-0.65	0.66	-0.13	0.97	-0.37
37.PDGFRA	R	-0.51	-0.07	-0.0013	1.49	-0.67

To ascertain if the apparent reduction in apoptotic function between human and chimpanzee brains extends to other tissues, we analyzed the log-ratio values of the 37 differentially expressed genes involved in the apoptotic pathway in heart, liver, kidney and testis (Table 2.3). The results indicate that the vast majority of the genes involved in apoptotic function that are differentially expressed in the human vs. the chimpanzee brain are similarly differentially expressed in the other organs (Table 2.3). For example, 11 of the 16 (69%) inducers of apoptosis that are down-regulated in the human brain are similarly down-regulated in at least three of the other four organs examined. All 9 (100%) repressors of apoptosis that are up-regulated in the human brain are similarly up-regulated in at least three of the four other organs examined. The overall correlation between the patterns of expression of the known inducers and repressors of apoptotic function between brain and the four other organs examined is highly significant (Table 2. 3, p-value = $6.15 * 10^{-11}$, Signs test; n+ = 82, n- = 18).

Consequences of the hypothesis and discussion

Since it is well known that reduced apoptotic function is associated with an increased risk of cancer and related diseases (Lowe and Lin 2000), our results are consistent with the hypothesis that selection for increased cognitive ability in humans may have coincidently resulted in an increased risk of cancer and other diseases associated with reduced apoptotic function. Consistent with our hypothesis is the recent observation that patients with human cognitive diseases associated with increased brain neuron apoptosis display significantly reduced rates of cancer (Driver et al. 2007; Eskenazi et al. 2007).

Given that the onset of cancer typically begins well after the onset of reproductive age

(http://seer.cancer.gov/csr/1975_2005/), the negative selective cost of reducing apoptotic function to increase cognitive ability within the human lineage was most likely minimal. Since there is no apparent advantage to systemically reducing apoptotic function, the alternative hypothesis that a reduced rate of neuron death in humans could have been an indirect consequence of selection for reduced apoptotic function in other organs is unlikely.

While differences in patterns of gene expression between humans and chimpanzees predict that apoptotic function is reduced in the human brain, additional laboratory studies as well sequence analysis (Table A.2 and Figure A.1) will be required to validate predictions from the gene expression studies. Moreover, as additional gene expression assays are conducted in other primates, it should be possible to test the hypothesis that reduced apoptotic function is a derived condition in the human lineage.

In summary, recent comparative gene expression results between chimpanzees and humans are consistent with the view that natural selection for increased cognitive ability in humans was attained, at least in part, by a reduction in the level of programmed cell death in the human brain. Existing data indicate that this reduction in apoptotic function extends beyond the brain to other tissues and is consistent with an evolutionary hypothesis whereby natural selection for increased brain size and cognitive ability in the human lineage coincidently resulted in an elevated propensity for cancer and other diseases associated with reduced apoptotic function.

ACKNOWLDEGEMENTS

We thank DeEtte Walker for editorial assistance. We are grateful to Dr.Phillip Khaitovich and Dr. Svante Pääbo for providing the Affymetrix .cel files used in our gene expression analysis.

CHAPTER 3

HUMAN CELLS DISPLAY A REDUCED APOPTOTIC FUNCTION RELATIVE TO THOSE OF CHIMPANZEES AND MACAQUES

ABSTRACT

Humans and chimpanzees differ in a number of phenotypic traits, among them being a larger sized human brain and an increased propensity for cancer in humans. Previously published gene expression analyses suggest that apoptotic function may be reduced in humans relative to chimpanzees and that this difference may account for both the relatively larger human brain and the increased propensity of humans to develop cancer. In this study, we directly test the hypothesis that humans maintain a reduced apoptotic function relative to chimpanzees by conducting a series of apoptotic function assays on human, chimpanzee and macaque primary fibroblastic cells. Human cells consistently displayed significantly reduced apoptotic function relative to the other primate cells. These results are consistent with earlier findings indicating that apoptotic function is reduced in humans relative to other primates and with the hypothesis that selection for increased cognition in humans may have indirectly resulted in an increased risk of cancer.

INTRODUCTION

Among the many well-documented phenotypic differences that exist between humans and chimpanzees is the exceptionally large size of the human brain with certain areas of specialized function (Carroll 2005; de Sousa and Wood 2007) and an increased incidence

of cancer among humans (Seibold 1973). The human brain is approximately three times as large as the chimpanzee brain, and epithelial neoplasms have been shown to be responsible for more than 20% of human deaths, compared to less than 2% in chimpanzees (Seibold 1973). We previously reported the results of gene expression and pathway data analyses indicating that humans and chimpanzees harbor significant differences in apoptotic function and that this difference might help explain both the relatively larger size of the human brain and the increased propensity of humans to develop cancer (Arora et al. 2009).

During neuronal development, apoptosis functions by adjusting the initial pool of neuronal progenitor cells for the proper morphogenesis of the nervous system. In a number of knockout studies conducted with mice, it has been shown that key pro-apoptotic factors, such as Apaf1 (Honarpour et al. 2001), caspase-9 (Joza et al. 2001) and caspase-3 (Kuida et al. 1996), play important roles in the development of the brain. In addition, elevated levels of apoptotic pathway genes have been associated with several human neurodegenerative diseases. For example, in both Parkinson's and Huntington's diseases, apoptotic genes have been shown to be significantly up-regulated (Tatton et al. 2003; Ferreira et al. 2010). Consistent with this observation, the pro-apoptotic factors caspase-1 and capsase-3 are significantly over-expressed in transgenic mouse models of Huntington's disease (Chen et al. 2000). These and other data implicating apoptosis in brain development suggest that the process may have been involved in human brain evolution as well. Indeed, comparative sequence analyses of apoptotic pathway genes

among primates suggest that the evolution of apoptotic genes has been accelerated within the human lineage relative to the other primates (Vallender and Lahn 2006).

Cancer, in contrast, is an example of a disease where the apoptotic pathway is typically suppressed resulting in over-proliferation and/or reduced destruction of tumor cells (Kerr et al. 1972). Indeed, reduced apoptotic function is now generally considered one of the hallmarks of cancer (Hanahan and Weinberg 2000).

Based on the above facts and our previous findings suggesting that apoptotic function may be reduced in humans relative to chimpanzees, we hypothesized that the selective pressure to increase cognitive functions/brain size in humans may have resulted in a reduced apoptotic function, having the indirect consequence of increasing the propensity of humans to develop cancer relative to chimpanzees (Arora et al. 2009). A keystone of this hypothesis is the assertion that apoptotic function is significantly reduced in humans relative to chimpanzees. While our analysis of gene expression differences between chimpanzees and humans is consistent with this assertion, we sought to acquire direct experimental support by monitoring relative apoptotic function in human, chimpanzee and macaque (out group) cell lines. The results are consistent with the assertion that apoptotic function is indeed reduced in humans relative to chimpanzees and macaques, and with the hypothesis that selection for increased cognitive ability within the human lineage may have indirectly increased the propensity of humans to develop cancer.

RESULTS

Selection of Cell lines

A number of previously published comparative studies between humans and non-human primates have successfully employed primary fibroblasts to study the differences between humans and non-human primates (e.g., gorilla and bonobo) (e.g., Karaman et al. 2003; Calarco et al. 2007). These previous studies have found that these fibroblastic cell lines display general molecular properties characteristic of the species. Thus, we chose to take a similar approach and conduct tests on primary fibroblast cell lines derived from humans, chimpanzees and macaques to determine relative apoptotic function among these species.

As an initial test of the suitability of using primary fibroblast cell lines to reflect differences in apoptotic function between these species, we compared the expression levels of apoptotic pathway genes that we previously reported to differ between human and chimpanzee tissues by gene expression analysis (Arora et al. 2009) with previously reported expression differences between primate fibroblast cell lines. While gene expression data derived from human fibroblast cells lines is available, gene expression profiling of chimpanzee fibroblast cells lines is not. However, such data is available for bonobo fibroblast cell lines. Bonobo (*Pan paniscus*) is a closely related species to chimpanzees (*Pan troglodytes*) and together they make up the genus *Pan* (Kaessmann et al. 1999). The two species are genetically nearly identical (99.4%). Thus, consistent with accepted standards in the field (e.g., Karaman et al. 2003), we considered the gene

expression patterns of the bonobo cell line to be representative of chimpanzee for the purpose of our initial analysis.

We compared the expression pattern of the 37 apoptotic genes that we previously identified as being differentially expressed between human and chimpanzee brains (Arora et al. 2009) (Table 3.1). Among the 21 genes identified as inducers of apoptosis in the brain dataset, 4 (19%) were not found to be expressed in the fibroblast dataset. Of the remaining 17, twelve genes in the fibroblast dataset (70%) displayed the same pattern of differential expression between human and chimpanzee/bonobo as those in the brain. Among the 16 genes identified as repressors of apoptosis in the brain dataset, one (6%) was not expressed in the fibroblasts while eight genes in the fibroblast dataset (53%) displayed the same pattern of differential expression as those in the brain. With respect to the patterns of expression of the known inducers and repressors of apoptotic function, there is an overall significant correlation between the brain dataset and the fibroblast dataset (Signs Test, p-value = 0.013). This correlation indicates that fibroblasts are a suitable model for testing the differences in apoptotic function between humans and chimpanzees.

Table 3.1: Comparison of the human-chimpanzee brain dataset with the human-bonobo/gorilla fibroblast data set with respect to the expression patterns of the apoptotic genes. With respect to the expression pattern of apoptotic genes, the human brain dataset (Arora et al. 2009) (relative to chimpanzees) is comparable to the human fibroblast dataset (Karaman et al. 2003) (relative to bonobos and gorillas) (Sign's test, p-value=0.013).

Gene Symbol	Expression Pattern in human brain dataset relative to chimpanzees	Expression Pattern in human fibroblast dataset relative to bonobos and gorillas
INDUCERS		
1.AIFM1	Down	Down
2.BID	Up	Down
3.CAPN2	Down	Down
4.CAPN3	Down	Down
5.CAPN5	Down	Down
6.CAPN7	Down	Not detected
7.CASP2	Down	Down
8.CASP6	Down	Not detected
9.CASP9	Down	Down
10.CYCS	Down	Down
11.DIABLO	Down	Not Detected
12.HTRA2	Down	Up
13.MKK7	Down	Down
14.ASK1	Up	Up
15.MAP4K4	Up	Down
16.ERK1	Down	Up
17.ERK2	Down	Down
18.JNK1	Up	Down
19.R-RAS	Down	Up
20.PRKCE	Down	Not detected
21.SPTAN1	Down	Down
REPRESSORS		
22.IAP	Up	Up
23.IKKA	Up	Up
24.DDR1	Up	Down
25.DDR2	Up	Up
26.ICAD	Down	Down
27.FGFR2	Up	Down
28.KRAS	Up	Down
29.MEK2	Up	Up
30.RAF1	Down	Not detected

Table 3.1 continued

31.NFκB	Down	Down
32. TEK	Up	Up
33.EGFR	Up	Up
34.IKBKB	Down	Down
35.PARP1	Down	Down
36.TYRO3	Down	Up
37.PDGFRA	Down	Up

Human cells display higher cell viability than chimpanzee and macaque cells after treatment with apoptotic-inducing agents

To test for differences in cell viability between the human (AG13153), chimpanzee (S006007) and macaque (AG07915) cells after induction of apoptosis, cells were treated with the apoptotic-inducing agents staurosporine and mitomycin C (MMC). After treatment, differences in cell viability between the different species were measured using a cell viability assay.

Staurosporine is a natural product isolated from *Streptomyces staurosporeus.* It induces apoptosis by inhibiting protein kinase C, a known inducer of the anti-apoptotic Bcl-2 (Deng et al. 2000). MMC is a natural product isolated from *Streptomyces caespitosus* (Szybalski and Iyer 1964) and is a known chemotherapeutic agent used in the treatment of a number of cancers (Tomasz 1995). MMC is a DNA cross-linking agent and damages DNA by cross-linking bases in the same or adjacent strands of DNA. This cross-linking eventually triggers a powerful apoptotic stimulus, including the activation of p53 (Seong et al. 2005).

Cell lines derived from each species were treated with MMC, in a dose dependent manner, and cell viability was measured 72 hours after treatment using the Cell Titer

Blue assay in a 96 well plate. The experiment was replicated 4 times. Treating the cells with MMC significantly reduced viability of the chimpanzee (S006007) and macaque (AG07915) cells at relatively low doses of the drug, while the human cells (AG13153) displayed reduced viability only at higher concentrations of the drug (50 μM) (Figure 3.1). Relative cell viability in the chimpanzee cells were reduced from 112% (0.01 μM) to 52% (1 μM), followed by a steady decrease for the subsequent doses with a cell viability of 32% at the highest dose (150 μM) of the drug. In the case of the macaque cells, relative cell viability was reduced from 87% (0.01 μM) to 58% (1 μM), followed by another drop to 15% (50 μM), which steadily dropped to 4% at the highest dose (150 μM) of the drug.

In the case of the human cells, cell viability decreased from 98% (0.01 μM) to 56% (50 μM), decreasing to 37% at the highest dose (150 μM) of the drug. The relative cell viability differences between the human and the chimpanzees were significant (student's t-test, $p<0.05$) at 1 μM, 5 μM, 10 μM and 50 μM of the MMC. Likewise, differences between the humans and the macaques were significant at all doses of the drug, except 0.01 μM of MMC. In all cases, cell viability was higher in the humans compared to chimpanzees and macaques.

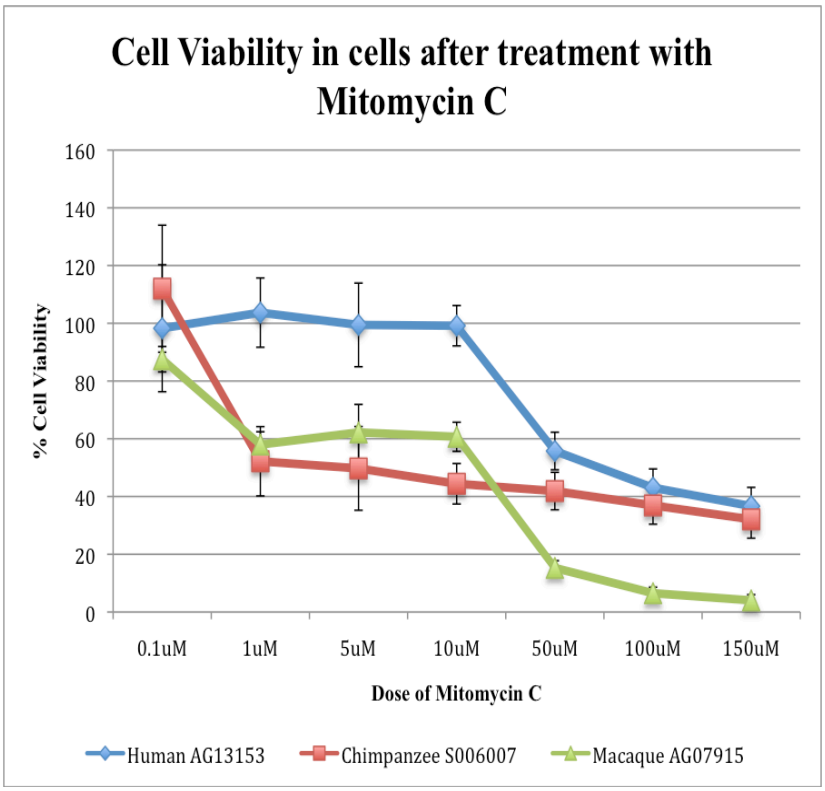


Figure 3.1: Relative viability of human (AG13153), chimpanzee (S006007) and macaque (AG07915) cells after treatment with mitomycin C (MMC). Cells were treated with different doses (0.1 μM – 150 μM) of the drug MMC for 72 hrs, after which cell viability was measured using the Cell Titer Blue assay. Each data point represents the mean of four replicates ± SD. Cell viability was expressed as a percentage of the viable cells in the treated group to the untreated control group (0 μM). The difference in cell viability was significant between humans and chimpanzees at 1 μM, 5 μM, 10 μM and 50 μM of MMC (student's t-test, $p<0.05$), and between humans and macaques at all doses except at 0.01 μM of MMC.

Similar results were seen when the human (AG13153) and chimpanzee (S006007) cells were treated with staurosporine (Figure 3.2), with the humans cells being more viable than the chimpanzee cells at various doses of staurosporine.

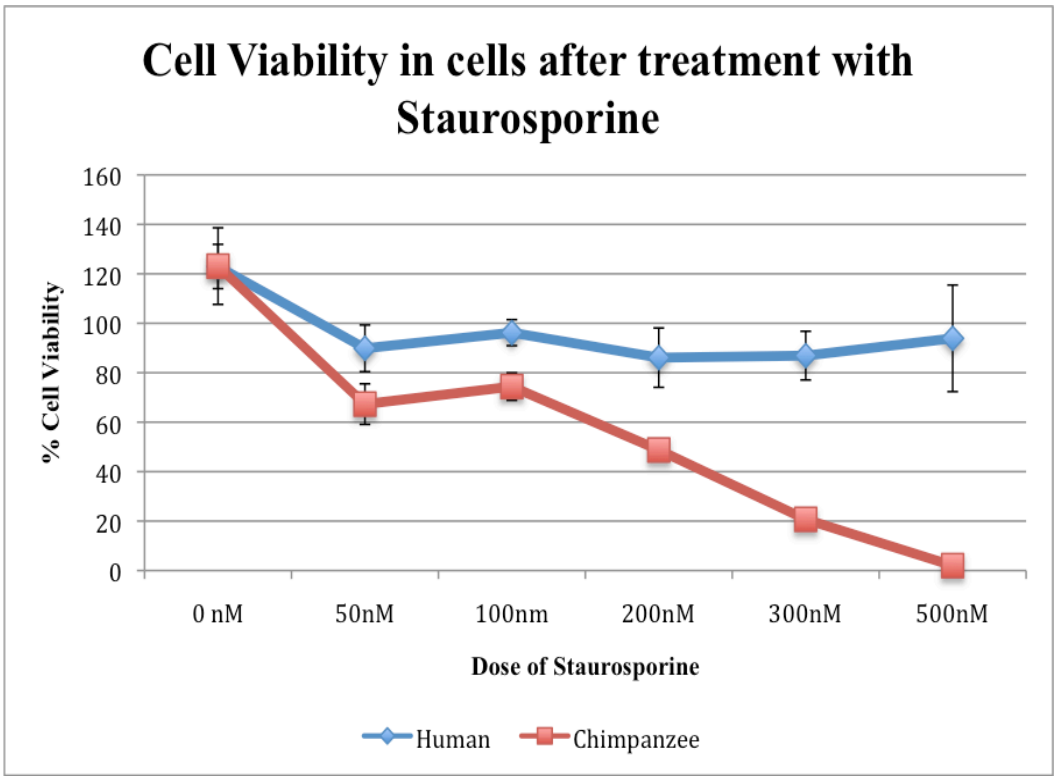


Figure 3.2: Relative cell viability between human (AG13153) and chimpanzee (S006007) cells after treatment with staurosporine. Cells were treated with different doses (0-500 nM) of the drug staurosporine for 48 hrs, after which cell viability was measured using the Cell Titer Blue assay. Each data point represents the mean ± SD of four replicates. Cell viability was expressed as a percentage of the viable cells in the treated group to the untreated control group (0 μM). The difference in cell viability was significant between humans and chimpanzees at all the doses of the drug (student's t-test, $p<0.05$).

Human cells treated with the apoptotic-inducing agent mitomycin C (MMC), display significantly higher IC_{50} values than chimpanzee or macaque cells

IC_{50} (half maximal inhibitory concentration) is a measure of the effectiveness of a compound or a drug to inhibit a biological function (e.g., cell viability) by 50% (Cheng and Prusoff 1973). The IC_{50} values of MMC were computed and compared between human (AG13153), chimpanzee (S006007) and macaque (AG07915) cell lines. The

results presented in Table 3.2, show that the MMC IC_{50} values are significantly higher (student's t-test, $p<0.05$) for the human cells than either chimpanzees or macaques, consistent with the hypothesis that human cells have reduced apoptotic function. The IC_{50} values in the chimpanzee and the macaque cell lines were 10.72 μM and 6.04 μM respectively, whereas in the human cell lines it was 78.87 μM.

Table 3.2: Relative IC_{50} values after treatment of human (AG13153), chimpanzee (S006007) and macaque (AG07915) cells with mitomycin C (MMC). The difference in the IC_{50} values between the humans and chimpanzees and between the human and macaques in both sets was significant (student's t-test, $p<0.05$)

Cell Line	IC_{50}
Human (AG13153)	78.87 μM
Chimpanzee (S006007)	10.72 μM
Macaque (AG07915)	6.04 μM

Human cells display phenotypic features characteristic of reduced apoptotic function relative to chimpanzee cells after treatment with the apoptotic-inducing agent mitomycin C (MMC)

Cells were treated with different doses of MMC (10μM, 15μM and 100μM), washed with PBS and stained with 10 μg/ml of Hoechst dye for 15 minutes. The cells were then visualized under a fluorescent microscope for characteristic features of apoptosis (Figure 3.3 A-H).

The nuclei of the control cells had an oval shape with homogeneous intensity (Figure 3.3A), whereas cells treated with MMC typically displayed phenotypic features that are characteristic of apoptosis including condensed and fragmented shapes with irregular staining homogeneity (Ziegler and Groscurth 2004). The phenotypic characteristics of the two human cell lines (AG13153 and AG07307) treated with 10 μM (Figure 3.3B) and 15μM (Figure 3.3E) of MMC respectively were similar to those of the control cells (Figure 3.3A). At a higher dose of MMC (100 μM), one human cell line (AG13153) showed phenotypic characteristics of apoptosis (Figure 3.3H).

The chimpanzee cells (S006007 and S005795) treated with 10 μM (Figure 3.3C) and 15 μM (Figure 3.3F) of MMC respectively showed DNA condensation and fragmentation consistent with onset of apoptosis. Likewise, there were similar results for the chimpanzee cells (S006007) treated with 100 μM of MMC (Figure 3.3I). Similar results were seen for the macaque cells (AG07915 and AG07128) treated with 10 μM (Figure 3.3D) and 15 μM (Figure 3.3G) of MMC respectively.

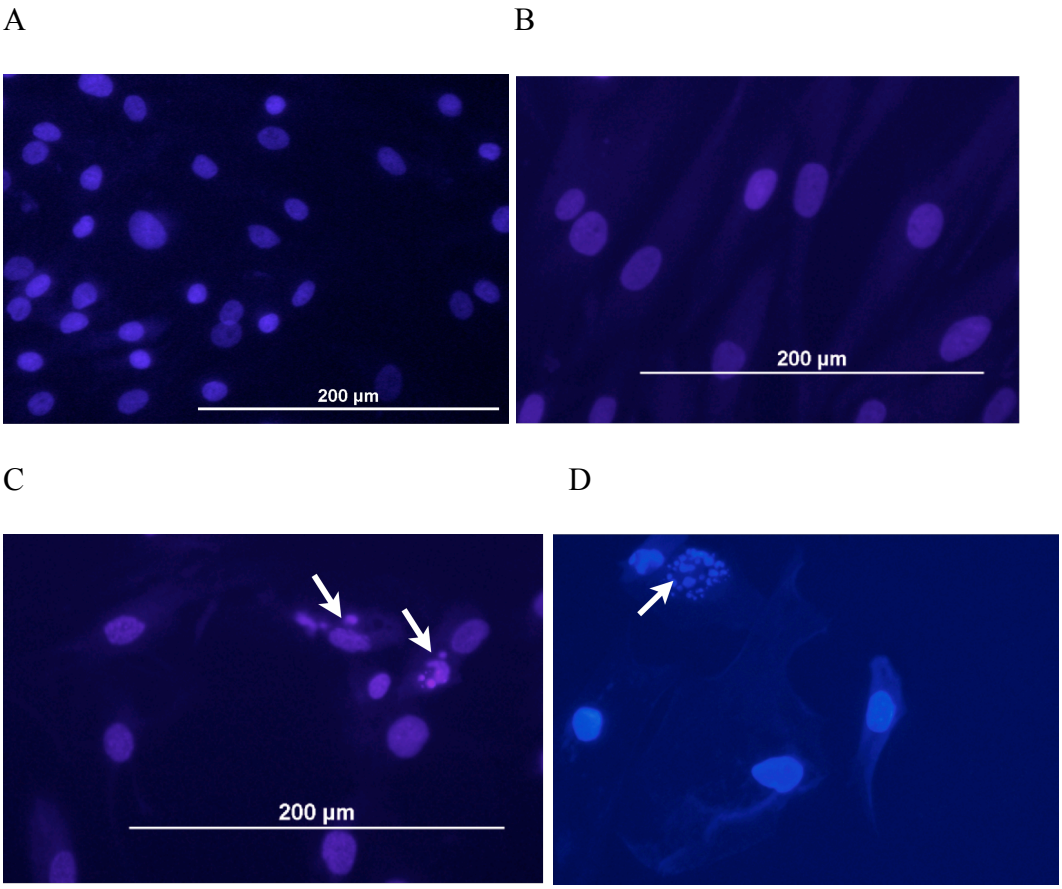
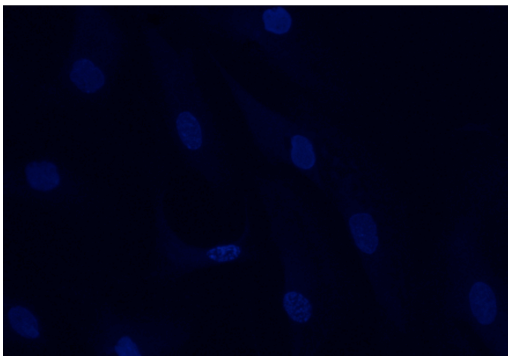
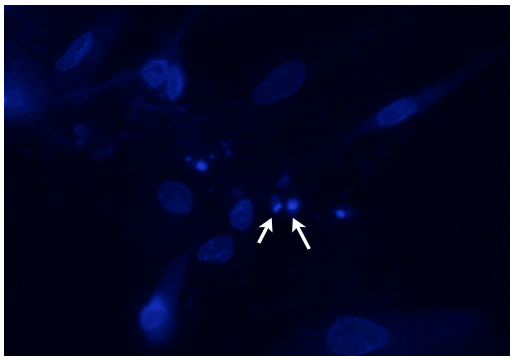


Figure 3.3: Human cells display morphological features characteristic of reduced apoptotic function relative to chimpanzee and macaque cells after treatment with mitomycin C (MMC). Cells were treated with the indicated concentration of MMC and then stained with 10 µg/ml of Hoechst and visualized under the fluorescence microscope. A) Chimpanzee cells (S006007) used as control; B) Human cells (AG13153) treated with 10 µM of MMC; C) Chimpanzee cells (S006007) treated with 10 µM of MMC; D) Macaque cells (AG07915) treated with 10 µM of MMC; E) Human cells (AG07307) treated with 15 µM of MMC; F) Chimpanzee cells (S005795) treated with 15 µM of MMC; G) Macaque cells (AG07128) treated with 15 µM of MMC; H) Human cells (AG13153) treated with 100 µM of MMC and I) Chimpanzee cells (S006007) treated with 100 µM of MMC. White arrows in the picture indicate cells showing phenotypic characteristics of apoptosis, such as DNA fragmentation (Figures 3.3C, D, F, G and H). The control (Figure 3.3A) showed normal phenotypic characteristics without any condensation or DNA fragmentation.

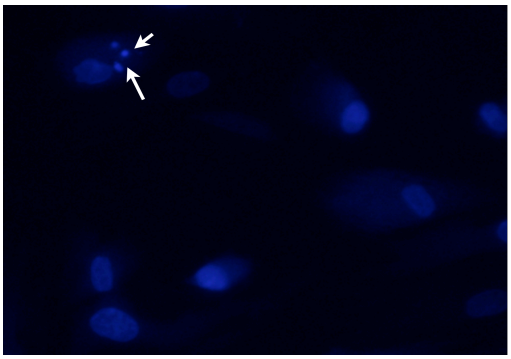
E



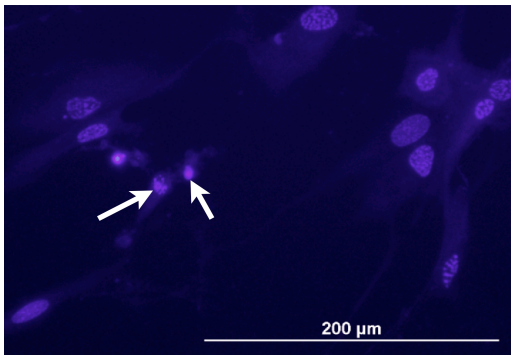
F



G



H



I

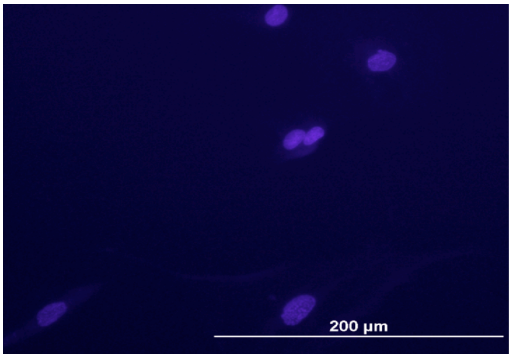


Figure 3.3 contd.

Human cells display lower caspase-3/7 activity than chimpanzee cells after treatment with staurosporine

The executioner caspase-3/7 is activated during apoptosis and is considered a biomarker of the process (Elmore 2007). Thus, the activity of these caspases was compared between human (AG13153) and chimpanzee (S006007) cells after treatment with increasing doses (100 nM and 300 nM) of the apoptotic-inducing drug staurosporine using the Caspase-3/7 Glo Assay (Liu et al. 2004). For both doses of the drug, the chimpanzee cells displayed significantly (student's t-test, $p<0.05$) higher caspase-3/7 activity than the human cells (Figure 3.4), consistent with the hypothesis that apoptotic function is reduced in human cells.

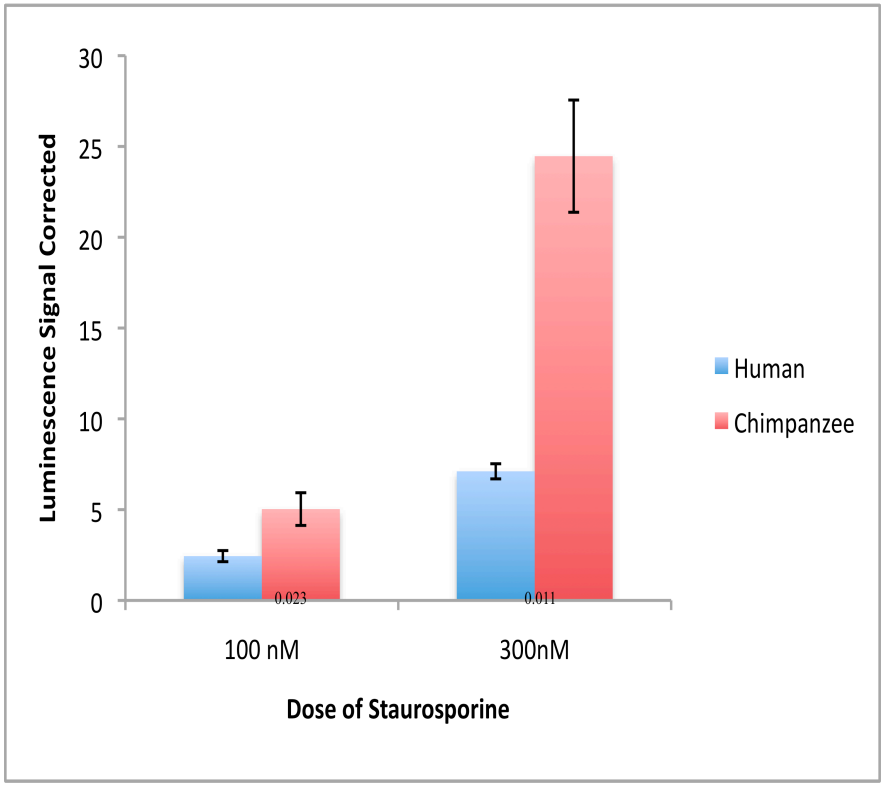


Figure 3.4: Human (AG13153) cells display lower caspase-3/7 activity than chimpanzee cells (S006007) after treatment with staurosporine.
Chimpanzee cells displayed significantly higher caspase-3/7 activity than human cells at 100 nM of staurosporine (student's t-test, p=0.023) and at 300 nM of staurosporine (student's t-test, p=0.011). The luminescence signal is proportional to the caspase-3/7 activity.

Human cells display reduced release of apoptotic factors relative to chimpanzee cells after treatment with the apoptotic-inducing drug mitomycin C (MMC)

The apoptotic pathway consists of the extrinsic and the intrinsic pathways (Elmore 2007), both of which converge on activating the executioner caspases, caspase-3 and caspase-7 (Tait and Green 2010). The intrinsic pathway proceeds through the release of pro-apoptotic factors (cytochrome C, Diablo and AIF) present in the mitochondrial intermembrane space, which then leads to the activation of apoptosis. This release of pro-

apoptotic factors occurs due to the permeabilization of the inner and outer mitochondrial membranes (Kroemer and Reed 2000).

Under normal physiological conditions, a transmembrane electrical potential gradient ($\Delta\Psi_m$) is maintained across the mitochondrial membranes, and this gradient is indicative of the normal functioning of the mitochondria (Mitchell 1961). On the induction of apoptosis, the mitochondrial outer membrane permeabilizes, resulting in the release of pro-apoptotic factors as well as a reduction of $\Delta\Psi_m$ (Budd et al. 2000). Reduction of $\Delta\Psi_m$ is indicative of apoptosis and can be easily measured using lipophilic, cationic fluorescent redistribution dyes, such as tetramethyl rhodamine ethyl (TMRE) and methyl (TMRM) (Ehrenberg et al. 1988).

In healthy, non-apoptotic cells, these lipophilic dyes bearing a delocalized positive charge enter the negatively charged mitochondria and fluoresce at a certain intensity. On the induction of apoptosis, as $\Delta\Psi_m$ is disrupted, these dyes no longer accumulate in the mitochondria and become dispersed in the cytosol, resulting in an overall drop in the level of cellular fluorescence (Rasola and Geuna 2001). This drop in fluorescence is indicative of apoptosis. Using flow cytometry, the number of cells undergoing this drop in fluorescence can be measured.

These dyes were used to test whether the number of apoptotic cells differed between humans and chimpanzees cell lines once apoptosis was induced. The human (AG07307) and chimpanzee (S005795) cell lines were treated with 30 μM and 100 μM of MMC.

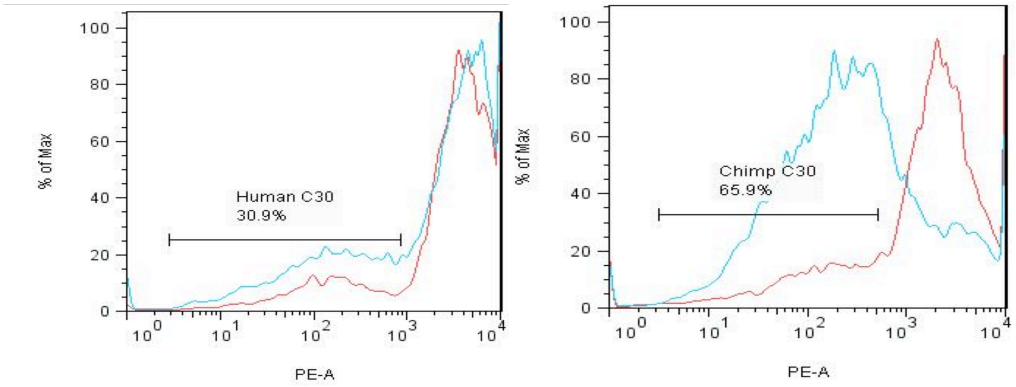
After 72 hours, the cells were treated with 100 nM of TMRE dye, and then the amount of apoptotic cells was measured in the two samples.

The chimpanzee cells, following treatment with 30 μM of MMC, had approximately 69.1% (average of two replicates) apoptotic cells, whereas the human cells treated with the same dose had approximately 33.8% apoptotic cells (Table 3.3 and Figure 3.5A). Similarly, the chimpanzee cells treated with 100 μM of MMC had 72.9% apoptotic cells, compared to the human cells, which had 42.8% apoptotic cells (Table 3.3 and Figure 3.5B). The difference in the number of apoptotic cells between the two samples was significant (proportions test, $p<0.05$) at both doses of the drug.

Table 3.3: Number of human (AG07307) and chimpanzee (S005795) cells undergoing apoptosis, expressed as a percentage of the total number of cells detected by flow cytometry, when treated with 30 and 100 μM of mitomycin C (MMC). Human (AG07307) and chimpanzee (S005795) cells were treated with 30 μM and 100 μM of the apoptosis-inducing agent mitomycin C for 72 hours. Subsequent to treatment for 72 hours, the percentage of viable cells remaining was monitored by flow cytometry using fluorescence the lipophilic dyes, tetramethyl rhodamine ethyl (TMRE) and methyl (TMRM). The results indicate that human cells are significantly (proportions test, $p<0.05$) more resistant to treatment by mitomycin C than chimpanzee cells reflecting reduced apoptotic function.

	Untreated	**30 μM MMC**	**100 μM MMC**
Human	5.15%	36.6%	42.8%
		30.9%	
Chimpanzee	7.8%	72.3%	72.9%
		65.9%	

3.5 A



3.5 B

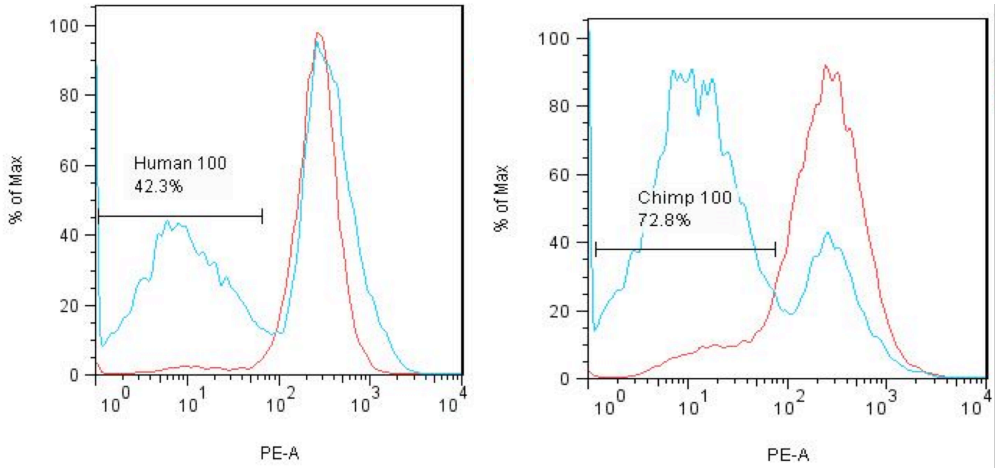


Figure 3.5: Mitochondrial membrane potential ($\Delta\Psi_m$) is reduced to a lesser extent in human cells (AG07307), relative to chimpanzee cells (S005795) when treated with A) 30 µM and B) 100 µM of mitomycin C (MMC). Human and chimpanzee cells were treated with different doses of MMC, and mitochondrial membrane potential was detected by monitoring relative levels the fluorescence of lipophilic dyes, tetramethyl rhodamine ethyl (TMRE) and methyl (TMRM). A reduction in fluorescence is indicative of onset of apoptosis. The results indicate that chimpanzee cells are associated with higher apoptotic function than human cells (Red lines = untreated cells; Blue lines = cells treated with the indicated levels of MMC). Mitochondrial membrane potential was monitored by measuring fluorescence at 574 nm (FL-2), as represented by the x-axis, whereas the y-axis represents the number of cells.

DISCUSSION

Among the documented differences between humans and chimpanzees are the relative large size of the human brain with certain areas of specialized function and the susceptibility of humans to diseases like cancer. Since apoptosis plays a role in both the development of the brain and disease progression to cancer, it is plausible that these differences may be attributed to differences in apoptotic function. Previous comparative gene expression analyses from our lab suggested that humans may have a reduced apoptotic function compared to chimpanzees (Arora et al. 2009).

In an effort to directly test the hypothesis of a reduced apoptotic function in humans, human and chimpanzee cells were treated with known apoptotic-inducing agents and the relative response of the cells monitored. A variety of tests were conducted and the results were uniformly consistent with the hypothesis that the apoptotic function in humans is reduced compared to chimpanzees. For example, at low doses of the apoptotic agent MMC (10 μM and 15 μM), chimpanzee cells displayed a lower number of viable cells (Figure 3.1) compared to the human samples. The phenotypic characteristics of the chimpanzee cells at these doses were typical of apoptosis showing nuclear condensation and DNA fragmentation (Figure 3.3C and 3.3F). In contrast, the phenotypic characteristics of the human cells at these low doses (Figure 3.3B and 3.3E) were similar to the control or untreated cells (Figure 3.3A), indicating that very little or no apoptosis was occurring in the human samples, as also indicated by higher cell viability in human cells. The comparison of the IC_{50} values of MMC in the human and chimpanzee cells confirmed that the chimpanzee cells were more sensitive to MMC induced apoptosis

(Table 3.2). These and the other results are consistent with the hypothesis that humans are associated with reduced apoptotic function relative to chimpanzees.

To determine whether the observed differences in apoptotic function between humans and chimpanzees was most likely to have occurred in the human or chimpanzee lineage, we used macaque as an out-group in our assays. The results consistently indicated that macaque cells behaved like chimpanzee cells in displaying higher sensitivity to apoptotic inducing agents relative to human cells. These findings are consistent with the hypothesis that the reduced apoptotic function associated with human cells is an evolutionarily derived condition occurring within the human lineage subsequent to the divergence of humans and chimps from a common ancestor ≈ 6 MYA (Goodman 1996).

Previous studies have demonstrated that apoptosis plays a significant role in mammalian neuronal development (Kuida et al. 1998; Honarpour et al. 2001; Joza et al. 2001). Thus, it is possible that selection for increased cognitive ability in humans may have involved changes in the apoptotic pathway. Consistent with this possibility is previously published findings indicating that apoptotic genes display accelerated rates of evolutionary change within the human lineage relative to the other primates (Vallender and Lahn 2006). Our previous analysis of gene expression differences between chimpanzees and humans independently suggested that apoptotic function in the brain and other tissues may be significantly reduced in humans relative to chimpanzees (Arora et al. 2009). These differences led us to postulate that selective pressure for increased cognitive ability within the human lineage may have, at least in part, resulted in reduced apoptotic

function. Since a reduced apoptotic function is a well known hallmark of cancer, we further hypothesized that selection for increased cognitive ability may have indirectly resulted in an increased risk of cancer in humans relative to chimpanzees. Since the onset of most cancers occurs well beyond reproductive age, there may have been little or no negative consequence of the reducing apoptotic function to increase brain size within the human lineage.

Like many evolutionary hypotheses, this model cannot be definitively substantiated. However, there is growing evidence that the propensity of many medical conditions in humans can be linked with traits or functions that were adaptive within our evolutionary past (Nesse et al. 2010). The results of the molecular studies reported here substantiate our earlier gene expression analyses indicating that apoptotic function is significantly reduced in humans relative to chimpanzees. These findings coupled with earlier studies linking apoptotic function with neural development and cancer, are consistent with the hypothesis that the evolution of increased cognitive function in humans may have contributed to our increased propensity to develop cancer.

METHODS

Fibroblast Cell Lines and Reagents

Primary human, chimpanzee and macaque fibroblast cell lines were obtained from Coriell cell repositories (Camden, NJ, USA) after they were matched for age and gender and after the generation gap between the species was taken into account (Table 3.4). All the

primary fibroblast cultures were initiated from explants of 2-mm-skin biopsies and cultured in a selective medium.

Table 3.4: Fibroblast cell lines used in the experiments

Human	Chimpanzee	Macaque
AG13153 (30 year old Male)	S006007 (22 year old Male)	AG07915 (12 year old Male)
AG07307 (40 year old female	S005795 (26 year old female)	AG07128 (11 year old female)

Mitomycin C was obtained from Sigma Aldrich, MO, USA, and staurosporine from Fisher Scientific, PA, USA.

Cell Culture

Primary fibroblasts were cultured in a rich medium (DMEM supplemented with essential/non essential amino acids, vitamins, antibiotics and 10%FBS) at 37°C in 5% CO_2. The growth rates for each of the cell lines were determined in 96 well plates using the Cell Titer Blue (Promega Corporation, WI, USA) cell viability assay. Passage numbers at which tests were done were as follows: human (AG13153) P14-P16; chimpanzee (S006007) P14-P16; macaque (AG07915) P16-P18; human (AG07307) P17-P19; chimpanzee (S005795) P18-P20; and macaque (AG07128) P18-P20.

Cell Viability Assay

The cell viability experiments were conducted in 96 well plates. The number of cells seeded per well was determined from the growth curves and was 80,000 cells/ml for the human cell lines, 120,000 cells/ml for the chimpanzee cell lines and 60,000 cells/ml for the macaque cell lines. The cells were allowed to grow for 24 hours, after which some of them were treated with different doses of staurosporine and MMC for 48 and 72 hours respectively. Some were untreated controls. The dilutions for the drugs were done by using the stock solution of the drug and then diluting it with RPMI + 5% FBS media.

After treatment with the drugs, the cells were treated with 20 μl of Cell Titer Blue reagent (Promega Corporation, WI, USA), followed by incubation at 37°C for two hours. The fluorescence was then measured in the range of $560_{EX}/590_{EM}$ in a fluorescence plate reader (Bio-Tek Multi detection microplate reader, Bio-Tek, VT, USA). The values obtained for both treated and untreated samples were averaged for each dose of the drug, and then cell viability was calculated as per manufacturer's instructions (Promega Corporation, WI, USA).

Hoechst Staining

The cells in the 96 well plates were washed twice with 100 μl of PBS, followed by staining with 10 μg/ml of Hoechst dye for 15 minutes. The cells were then visualized using a microscope (Olympus IX51, Olympus, NJ, USA) and photographed using an Olympus DP72 digital camera.

Caspase-3/7 Activity Assay

The human (AG13153) and chimpanzee (S006007) cells in the 96 well plates were allowed to grow for 24 hours, followed by treatment with staurosporine for 48 hours to induce apoptosis. Following treatment with the drug, 100 μl of the Caspase-3/7 Glo reagent (Promega Corporation, WI, USA) was added, followed by incubation at 37°C for 30 minutes to generate a luminescence signal. The caspase-3/7 activities were determined by monitoring the activity on a luminescence plate reader (Spectramax Gemini XS Microplate Spectrofluorometer, Molecular Devices, CA, USA).

Measurement of Mitochondrial Transmembrane Potential

Changes in $\Delta\Psi_m$ were detected using tetramethyl-rhodamine ethyl (TMRE) or methyl (TMRM). Human (AG07307) and chimpanzee (S005795) cells were grown in 10 ml Petri dishes for 24 hours, after which they were treated with different doses of MMC for 72 hours. Trypsinized cells were combined with the supernatant medium and incubated with 100 nM of TMRE/TMRM for 30 minutes in the dark at 37°C. TMRE fluorescence was measured using the FL2 (574 nm) of the BD LSR II flow cytometer. Analysis of the data was done using Flo Jo 7.6 software.

Statistical Analysis

To analyze the differences in cell viability and caspase-3/7 activities, a student's t-test was performed in Excel. The proportions test, done in R, was used to analyze differences in the number of apoptotic cells between humans and chimpanzees. Mean differences were considered to be significant when $p<0.05$.

Analyses of IC_{50} values were computed from concentration-response curves using GraphPad Prism 5.0 (Graph Pad software, San Diego CA, USA).

CHAPTER 4

INDEL VARIATION BETWEEN HUMANS AND CHIMPANZEES IS CORRELATED WITH DIFFERENCES IN GENE EXPRESSION*

ABSTRACT

Background

Although humans and chimpanzees have accumulated significant differences in a number of phenotypic traits since diverging from a common ancestor about six million years ago, their genomes are >98.5% identical at the protein coding loci. This modest degree of nucleotide divergence is not sufficient to explain the extensive phenotypic differences between the two species. It has been hypothesized that the genetic basis of the phenotypic differences lies at the level of gene regulation, which is possibly associated with the extensive INDEL (insertion/deletion) variation between the two species.

To test the hypothesis that large INDELs (80-12,000 bp) may have contributed significantly to differences in gene regulation between the two species, we have categorized human-chimpanzee INDEL variation mapping in or around genes and determined whether this variation is significantly correlated with species differences in gene expression.

* This chapter reports on a collaborative study, in which my contribution focused on the gene expression differences.

Results

We found that there is a significant correlation between differences in gene expression and INDEL variation between chimpanzees and humans predominantly involving the insertion of interspersed (predominantly retrotransposon) and non-interspersed sequences in the human lineage. The majority of this functionally significant INDEL variation was mapped to introns and arose in the human lineage after humans and chimpanzees diverged from a common ancestor.

Conclusion

Our results are consistent with the hypothesis that INDELs have played a significant role in human/chimpanzee evolution. Most of the differential gene expression seen between the two species may have been driven by accelerated regulatory evolution in the human lineage.

BACKGROUND

Although humans and chimpanzees have accumulated significant differences in a number of phenotypic traits since diverging from a common ancestor about six million years ago, their genomes are >98.5% identical at the protein coding loci (Mikkelsen et al. 2005). This modest degree of nucleotide divergence does not seem sufficient to explain the extensive phenotypic differences between the two species, and it has been hypothesized that the genetic basis of the differences lies at the level of gene regulation (King and Wilson 1975), which is possibly associated with the extensive insertion/deletion (INDEL) variation existing between the two species (Britten 2002).

A number of comparative genomic studies carried out between humans and non-human primates have revealed that significant INDEL variation exists between these species (Frazer et al. 2003; Watanabe et al. 2004; Chen et al. 2007a; Chen et al. 2007b). In one of the first studies done to identify human and chimpanzee INDEL variation, a 27 Mb region of the human chromosome 21 was compared with chimpanzee DNA sequences, and approximately 57 INDELs were identified (Frazer et al. 2003). In another study comparing the human chromosome 21 with the syntenic chimpanzee chromosome 22, as many as 68,000 INDELs were identified (Watanabe et al. 2004). In a recent comparison of human chromosome 21 and chimpanzee chromosome 22, as many as 6279 INDELs of size >10 bp were identified (Volfovsky et al. 2009).

We have previously shown that interspersed repeats, particularly retrotransposons, have contributed significantly to the INDEL variation between humans and chimpanzees (Polavarapu et al. 2006). Similarly, in another study done to classify recently inserted retrotransposons between humans and chimpanzees, it was shown that humans have a greater rate of insertions compared to chimpanzees, and the majority of these insertions are Alu elements and is located in or around genes (Mills et al. 2006b). Since retrotransposon sequences located in or around genes have the capacity to significantly alter patterns of gene expression, it has long been recognized that these sequences may be important factors in regulatory evolution (McClintock 1984; McDonald 1993; Mills et al. 2006a).

Another source of INDEL variation between humans and chimpanzees is simple tandem repeats (Tautz et al. 1986; Du et al. 1997). It has been proposed that tandem repeats located in or around genes alter gene expression by interacting with transcription factors, by altering chromatin structure or by acting as potential protein binding sites (Hamada et al. 1984; Pardue et al. 1987; Yee et al. 1991). Since tandem repeats in or around genes are capable of altering gene expression, they have also been postulated to play a role in regulatory evolution (Pardue et al. 1987; Sinha and Siggia 2005; Tomilin 2008), e.g., it was shown that variation associated with tandem repeats in the regulatory region of the vasopressin 1a receptor gene may account for the differences in social behavior between humans and chimpanzees (Hammock and Young 2005).

In this study, we present a detailed characterization of large INDEL variation (80-12,000 bp in length) associated with human and chimpanzee genes and test if this variation is significantly correlated with differences in gene expression in a variety of tissues. Our results indicate that both interspersed repeats (predominantly retrotransposons) and non-interspersed sequences, mapping to introns, have contributed significantly to human/chimpanzee regulatory evolution, primarily due to insertions within the human lineage.

RESULTS

Characterization of human and chimpanzee gaps

We use the terms "human gaps" (HGs) to refer to sequences present in chimpanzees but absent in humans and "chimpanzee gaps" (CGs) to sequences present in humans but absent in chimpanzees (Polavarapu et al. 2006). Collectively, these gaps constitute the INDEL variation between humans and chimpanzees. Using the database available at the UCSC Genome Bioinformatics web site (Kent et al. 2002), a total of 11365 HGs and 15144 CGs were identified (Table 4.1). The majority of these gaps (63 % of HGs and 75% of CGs) are associated with interspersed repeats, i.e. sequences that are repeated multiple times throughout the genome, and nearly all of these (>99%) are homologous to retrotransposon sequences (REs). Of the remaining gaps (37% of HGs and 25% of CGs), comprised of non-interspersed sequences (NISs), 33% are made up of tandem repeats and 67% are small INDELs of unique sequence (Table 4.1).

The presence of a sequence in humans (or *vice versa* in chimpanzees) that is missing at an orthologous genomic position in chimpanzees (humans) can either be due to an insertion in one species or a deletion in the other. By using macaques (*Macaca mulatta*) as an out-group (Figure 4.1), it was determined that the majority of the INDEL variation between humans and chimpanzees is due to insertions. For example, 57% of the retrotransposon-associated human INDELs are the result of chimpanzee insertions (CIs), and 43% are the result of human deletions (HDs), while 76% of the retrotransposon-associated chimpanzee INDELs are due to human insertions (HIs), and 24% are due to

chimpanzee deletions (CDs) (Table 4.2a). While the majority of non-interspersed sequences associated with tandem repeats (57% of the human INDELs and 61% of the chimpanzee INDELs) are insertions, the majority of the unique non-interspersed sequences (60% of the human INDELs and 53% of the chimpanzee INDELs) are due to deletions (Table 4.2b).

Table 4.1: Number of INDELs associated with different categories of sequences. Human (HGs) and chimpanzee gaps (CGs) constitute the INDEL variation between the two species. The INDELs contain interspersed repeats and non-interspersed sequences. Interspersed repeats are transposable element sequences that are present multiple times throughout the genome. The majority of interspersed repeats are retrotransposon sequences (sub categories: SINEs, LINEs, ERVs, SVAs and MEs. See text for further characterization). DNA transposable elements constitute <1% of interspersed repeats. Non-interspersed sequences (NISs) are tandem repeats or unique sequences that map to specific (INDEL) sites in the genome.

Categories of Gaps	Human Gaps (HGs)	Chimpanzee Gaps (CGs)
Total number of gaps	11365	15144
Interspersed Repeats (all)	7176	11398
Interspersed sequences (retrotransposons)	7121	11355
Retrotransposons (SINEs)	3494	7021
Retrotransposons (LINEs)	1847	2052
Retrotranspsons (ERVs)	519	356
Retrotransposons (SVAs)	114	681
Retrotransposons (MEs)	1147	1245
Interspersed sequences (DNA elements)	55	43
Non-interspersed sequences (all)	4189	3746
Non-interspersed sequences (tandem repeats)	1266	1334
Non-interspersed sequences (unique sequences)	2923	2412

Table 4.2: Number of human and chimpanzee INDELs associated with a) retrotransposons and b) non-interspersed sequences. Using macaque as an out-group, INDEL variation was characterized as chimpanzee insertions (CIs), human deletions (HDs), human insertions (HIs) and chimpanzee deletions (CDs).

a)

	Human INDELs			Chimpanzee INDELs		
Retrotransposon subclass	**CIs**	**HDs**	**CIs + HDs**	**HIs**	**CDs**	**HIs + CDs**
SINE	2264	1230	3494	5787	1234	7021
LINE	1311	536	1847	1756	296	2052
ERV	208	311	519	156	200	356
SVA	98	16	114	680	1	681
ME	154	993	1147	269	976	1245
Total	4035	3086	7121	8648	2707	11355

b)

	Human INDELs			Chimpanzee INDELs		
Non-interspersed sequence subclass	**CIs**	**HDs**	**CIs + HDs**	**HIs**	**CDs**	**HIs + CDs**
Tandem repeats	720	546	1266	814	520	1334
Unique sequences	1156	1767	2923	1145	1267	2412
Total	1876	2313	4189	1959	1787	3746

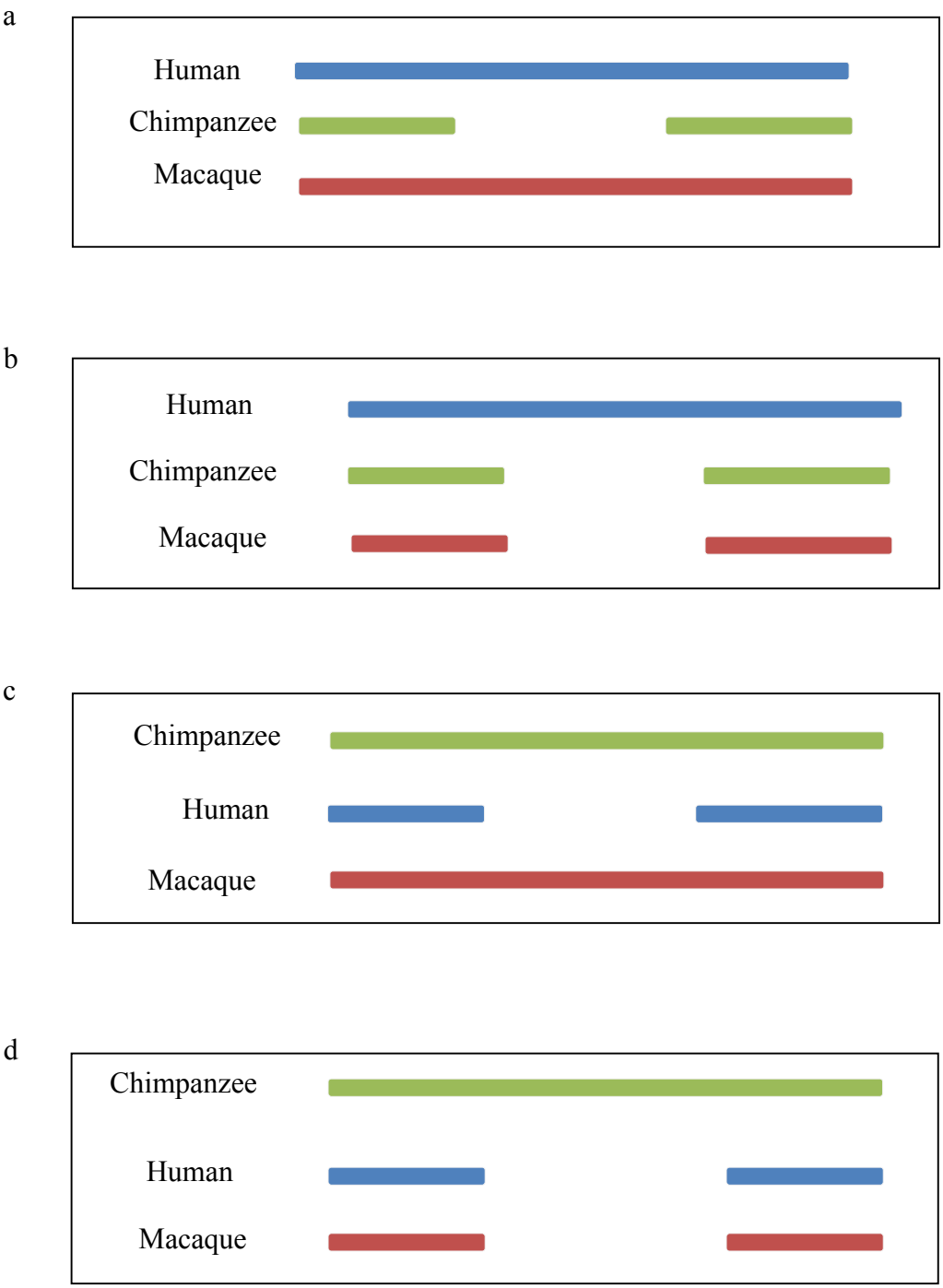


Figure 4.1: Method of characterizing INDELs as insertions or deletions in the chimpanzee or human lineage. (a) A region of the chimpanzee chromosome (green) showing a gap when it is aligned against a homologous region of the human chromosome (blue). The sequence corresponding to the gap region in the chimpanzee chromosome is found in the macaque (out-group) chromosome (red). The presence of the sequence in both the human and macaque chromosomes, but its absence in the chimpanzee

chromosome characterizes the gap as a chimpanzee deletion (CD); analogous reasoning was used to identify (b) human insertions (HIs); (c) human deletions (HDs) and (d) chimpanzee insertions (CIs).

Analysis of interspersed repeats homologous to retrotransposon sequences

As mentioned above, nearly all of the interspersed repeats associated with HGs and CGs (~99%) are homologous to retrotransposon sequences (REs). The REs were grouped into five different classes: 1) SINEs (short interspersed nuclear elements), 2) LINEs (long interspersed nuclear elements), 3) ERVs (endogenous retroviruses), 4) SVAs (biologically active composite elements consisting of fragments of SINEs, VNTRs (variable number of tandem repeats), and Alu elements) and 5) MEs (mosaic elements, which are inactive sequences comprised of a mosaic of more than one class of the above retrotransposon homologous sequences). Of the REs associated with HGs, 49 % are homologous to SINEs, 26% to LINEs, 7% to ERVs, 2% to SVAs and 16% to MEs (Table 4.1). Of the REs associated with CGs, 62% are homologous to SINEs, 18% to LINEs, 3% to ERVs, 6% to SVA and 11% to MEs (Table 4.1). These values are proportional to the relative frequency of the various classes of retrotransposons in the human and chimpanzee genomes (Lander et al. 2001; Mikkelsen et al. 2005).

Consistent with the relative transpositional activity of retrotransposon families in humans and chimpanzees (Lander et al. 2001; Mikkelsen et al. 2005), it was seen that the majority of the insertions involve SINEs and LINEs (Table 4.2a). The frequency of ERV insertions is 1.3 fold higher in chimpanzees than in humans, predominantly due to the expansion of two chimpanzee-specific ERV families (CERV 1/PTERV1 and CERV 2) 3-

5 million years ago (Yohn et al. 2005; Maksakova et al. 2006; Polavarapu et al. 2006). In contrast, the frequency of SVA insertions in humans is > 6-fold higher than in chimpanzees (Table 4.2a). The overall frequency of REs insertions is >2 fold higher in humans than in chimpanzees. The frequency of LINE, SVA and ERV deletions is higher in humans than in chimpanzees, while the frequency of SINE and ME deletions are nearly the same in both species (Table 4.2a).

Lineage specific interspersed (retrotransposon) insertions are responsible for the majority of INDELs located in or around human and chimpanzee genes

Of the 34,914 human/chimpanzee genes listed in the ENSEMBL database (Hubbard et al. 2007), 10597 or 30% (10597/34,914) were found to be associated with INDELs, i.e., having one or more INDELs located in or within 5 kb upstream or downstream of genes (Table 4.3). There are 6873 genes associated with INDELs containing REs only; 2908 genes associated with INDELs containing NISs only; and 816 INDEL-associated genes containing both REs and NISs. The number of genes with INDELS containing RE sequences (6873 + 816 = 7689) is greater than the number of genes associated with INDELs containing NISs (2908 + 816 = 3724).

A greater number of the genes associated with INDELs containing REs were classified as HIs (3149 + 326 = 3475). The number of genes associated with INDELs containing NISs and classified as HIs (718 + 326 = 1044) was greater than the number of genes associated with INDELs containing NISs and classified as HDs (740 + 155 = 895). In summary, the majority of the human-chimpanzee INDEL variation located in or around human and

chimpanzee genes is due to the insertion of REs in the human lineage after divergence from the human-chimpanzee common ancestor > 6 MYA (million years ago).

Table 4.3: Number of genes associated with INDELs. The genes associated with INDELs were classified depending on the type of INDEL (HIs, CIs, HDs, and CIs) associated with the gene and type of sequence contained in the INDEL (REs vs. NISs).

Type of INDEL	Genes associated with INDELS containing REs only	Genes associated with INDELs containing NISs only	Genes associated with INDELs containing both REs & NISs	Total (genes associated with INDELs)
HI	3149	718	326	4193
CI	1276	674	175	2125
HD	1139	740	155	2034
CD	1309	776	160	2245
Total	6873	2908	816	10597

INDEL variation is correlated with differences in gene expression

To explore the relationship between human-chimpanzee INDEL variations and the differences in gene expression, a previously published human-chimpanzee expression dataset consisting of expression arrays from five different tissues (brain, testis heart, liver and kidney) (Khaitovich et al. 2005) was reanalyzed. A major goal of this previous study was to correlate sequence differences with expression differences, and a number of microarray probe sets for which quality sequences could not be obtained in humans and chimpanzees (e.g., required for the calculation of ka/ki values) was excluded. Since the quality of the chimpanzee genome sequence has improved in recent years, and since our interest is in the possible contribution of INDELs to chimpanzee–human expression differences, this microarray dataset was reanalyzed by including the probe sets that were

previously excluded.

The most dramatic difference in gene expression between humans and chimpanzees was seen in the testis (70% of genes display a significant difference in expression between chimpanzees and humans), followed by heart (51%), brain (49%), kidney (47%) and liver (39%) (Table 4.4).

Table 4.4: Number of genes differentially expressed between humans and chimpanzees across five tissues. The percentage in parenthesis is calculated by dividing the number of genes that were differentially expressed or not in each tissue by the total number of genes detected in that tissue.

	Brain	**Testis**	**Heart**	**Liver**	**Kidney**
Number of genes detected	14133	15445	13497	13684	14059
Number of genes differentially expressed	6884 (48.7%)	10803(69.9%)	6843 (50.7%)	5308(38.8%)	6589 (46.9%)
Number of genes not differentially expressed	7249 (51.3%)	4642 (30.1%)	6654 (49.3%)	8376 (61.2%)	7470 (53.1%)

As mentioned above, the majority of INDELs are homologous to REs (Table 4.3). The expression levels of genes associated with the various classes of REs (SINEs, LINEs, ERVs, SVAs and MEs) were analyzed to determine if any particular class is preferentially associated with genes displaying significant differences in expression. Genes associated with INDELs containing SINEs were differentially expressed significantly between humans and chimpanzees in testis (proportions test, p=8.2E-06), brain (proportions test, p=8.4E-04), kidney (proportions test, p=3.05E-06), heart

(proportions test, p=0.001) and liver (proportions test, p=0.03). Genes associated with INDELs containing other classes of REs, on the other hand, were differentially expressed significantly in relatively fewer tissues (Table 4.5). SINEs are the most abundant and transpositionally active class of retrotransposons in humans and chimpanzees and have frequently been associated with functionally important insertion mutations in humans (Deininger and Batzer 1999; Hasler and Strub 2006).

Table 4.5: Association of INDEL variation containing retrotransposon sequences (REs) with differential gene expression. The INDEL-associated genes were classified on the type of REs (SINE, LINE, ER, SVA and ME) present in the INDEL. The proportions test was used to associate differential gene expression between humans and chimpanzees with the INDEL variation by comparing the proportion of INDEL-associated genes that were differentially expressed (DE) with the proportion of INDEL-associated genes that were detected but not differentially expressed (Exp) between the two species. Genes associated with INDELs that contained SINEs showed significant (** $p<0.01$, * $p<0.05$) association with differential gene expression in all tissues, whereas genes associated with INDELs containing other type of RE sequences showed significant association with differential expression in relatively fewer tissues.

Type of RE sequence	**Brain**		**Testis**		**Heart**		**Liver**		**Kidney**	
	Exp	**DE**	**Exp**	**DE**	**Exp**	**DE**	**Exp**	**DE**	**Exp**	**DE**
SINE	1338	1425**	773	2131**	1213	1401**	1534	1051*	1315	1365**
LINE	415	431	259	617	362	409	456	302	421	369
ERV	57	71	42	90	46	66	56	63*	63	63
SVA	137	121	68	210*	111	128	155	86	131	122
ME	406	455*	256	633	361	423	457	310	374	429**

To determine more precisely the nature of the association of differential gene expression with INDEL variation, the differentially expressed genes were grouped into four categories with respect to the intra-genic location of the associated INDELs: 1) Exon, 2)

Upstream (within 5000 bp upstream of transcription start site), 3) Downstream (within 5000 bp downstream of transcription termination site), and 4) Intron. In addition, the differentially expressed genes were also grouped with respect to the composition of the INDEL (REs vs. NISs).

There was no significant association of differential gene expression with INDEL variation located in the region upstream of the transcription start site in any of the tissues (Table 4.6). There was association of differential gene expression with INDELs located in the exons in the liver only, where INDELs containing NISs were significantly associated with differential gene expression (proportions test, $p=0.002$). INDEL variation mapping downstream of the transcriptional termination site was seen to be significantly associated with differences in gene expression only in the kidney, particularly with INDELs containing REs (proportions test, $p= 6.5E-04$). In all the tissues, there was significant association (proportions test, $p<0.05$) of INDEL variation located in the introns with differential gene expression (Table 4.6).

Table 4.6: Association of INDEL variation with differential gene expression, based on the location and composition of the INDEL. Genes associated with INDELs were classified with respect to the intra-genic location of the INDEL (Exon, Upstream, Downstream and Intron) and composition of the INDEL – retrotransposon sequences (REs) vs. non-interspersed sequences (NISs). The proportions test was used to associate differential gene expression between humans and chimpanzees with INDEL variation by comparing the proportion of INDEL-associated genes that were differentially expressed (DE) with the proportion of INDEL-associated genes that were detected but not differentially expressed (Exp) between the two species. Genes associated with INDELs located in the intronic region showed significant (** $p<0.01$, *$p<0.05$) association with differential gene expression in all tissues, whereas genes associated with INDELs in the exon and downstream region showed significant association with differential expression in the liver and kidney respectively.

	Exon						**UpStream**						**DownStream**						**Intron**					
	All INDELs		REs		NISs		All INDELs		REs		NISs		All INDELs		REs		NISs		All gaps		REs		NISs	
	Exp	DE	Exp	DE	Exp	DE	Exp	DE	Exp	DE	Exp	DE	Exp	DE	Exp	DE	Exp	DE	Exp	DE	Exp	DE	Exp	DE
Brain	74	49	16	20	58	31	312	342	236	262	80	82	330	347	235	247	98	109	1803	1907**	1514	1624**	637	674**
Testis	31	102	9	31	22	74	196	520	148	397	50	130	198	543*	142	387	60	165	1037	2874**	893	2419**	365	984*
Heart	50	63	14	21	39	43	309	326	240	248	76	79	303	356	223	251	85	111	1602	1876**	1356	1598**	539	645**
Liver	50	57*	18	17	34	41*	403	250	312	183	96	70	384	283	275	200	114	90	2026	1413**	1722	1199**	685	479
Kidney	54	66	13	23	43	46	336	321	252	248	89	75	316	361**	226	262**	93	108	1787	1812**	1498	1545**	640	602

Genes differentially expressed in all tissues are preferentially associated with INDELs

If the presence of an INDEL around human and chimpanzee genes is not a significant contributing factor to differences in gene expression between the two species, the proportion of genes associated with INDELs should be approximately equal for differentially expressed and non-differentially expressed genes. This was not found to be the case in any of the tissues, and it was seen that the proportion of differentially expressed genes associated with INDELs was significantly greater (proportions test, $p<0.05$) than the proportion of non-differentially expressed genes associated with INDELs (Table 4.7).

In order to test whether INDELs containing REs or NISs were a significant contributing factor to differences in gene expression between humans and chimpanzees, the proportion of differentially expressed genes associated with INDELs containing REs or NISs was compared with their non-differentially expressed counterparts. The proportion of differentially expressed genes associated with INDELs containing REs was significantly greater (proportions test, $p<0.05$) than their non-differentially expressed counterparts in all tissues (Table 4.8a), whereas the differentially expressed genes associated with INDELs containing NISs was significantly greater (proportions test, $p<0.05$) than their non-differentially expressed counterparts in all tissues, except kidney (Table 4.8b). While our results are consistent with the hypothesis that INDELs contribute significantly to differences in genes expression between humans and chimpanzees, this is not to say that other types of mutations are functionally insignificant. Indeed, there are a substantial

number of genes not associated with INDELs that are also significantly differentially expressed not only in kidney and liver but also in all tissues (Table B.1).

Table 4.7: Association of INDEL variation with differential gene expression. All INDEL variation seen in each of the tissues was tested to see for association with differential gene expression between humans and chimpanzees. The proportion of differentially expressed genes (DE) associated with INDEL variation was significantly greater (proportions test, $p<0.05$) than the proportion of non-differentially expressed genes (Exp) associated with INDEL variation in all tissues. Proportions are given in parentheses.

Tissue	DE genes with INDELs / Total DE genes	Exp genes with INDELs/ Total Exp genes	Proportions Test (p-value)
Brain	2266/6884 (0.33)	2153/7249(0.30)	4.054E-05
Testis	3438/10803 (0.32)	1256/4642 (0.27)	3.93E-09
Heart	2233/6843 (0.33)	1948/6654 (0.29)	2.7E-05
Liver	1696/5308 (0.32)	2466/8376 (0.29)	0.0019
Kidney	2179/6589 (0.33)	2144/7470 (0.29)	2.35E-08

Table 4.8: Association of differential gene expression with INDEL variation containing a) retrotransposon sequences (REs) and b) non-interspersed sequences (NISs). INDEL variation containing REs and NISs seen in each of the tissues was tested to see for association with differential gene expression between humans and chimpanzees. The proportion of differentially expressed genes associated with INDEL variation was significantly greater (proportions test, $p<0.05$) than the proportion of non-differentially expressed genes associated with INDEL variation. Proportions are given in parentheses.

a)

Tissue	DE genes with INDELS containing REs/ Total DE genes	Exp genes with INDELs containing REs/ Total Exp genes	Proportions Test (p-value)
Brain	1916/6884 (0.28)	1790/7249 (0.25)	2.42E-05
Testis	2862/10803 (0.26)	1072/4642 (0.23)	9.63E-06
Heart	1876/6843 (0.27)	1636/6654 (0.25)	0.00019
Liver	1416/5308 (0.26)	2072/8376 (0.25)	0.012
Kidney	1843/6589 (0.28)	1776/7470 (0.24)	1.52E-08

Table 4.8 b)

Tissue	DE genes with INDELs containing NISs/ Total DE genes	Exp genes with INDELs containing NISs/ Total Exp genes	Proportions Test (p-value)
Brain	801/6884 (0.12)	762/7249 (0.1)	0.036
Testis	1193/10803 (0.11)	440/4642 (0.094)	0.0041
Heart	777/6843 (0.11)	658/6654 (0.098)	0.006
Liver	590/5308 (0.11)	838/8376 (0.1)	0.041
Kidney	732/6589 (0.11)	768/7470 (0.1)	0.11

DISCUSSION

Over the ~ 6 million years that the human and chimpanzee lineages have diverged from a common ancestor, the two species evolved a variety of distinctive morphological, behavioral, cognitive and other phenotypic traits (Varki and Altheide 2005). To explore the genetic basis of the phenotypic differences that distinguish humans from chimpanzees, a number of comparative genomic studies have been conducted in recent years (Li et al. 2001; Mikkelsen et al. 2005). Perhaps the most surprising finding coming out of these studies is the paucity of protein coding nucleotide variation existing between these two species, supporting earlier contentions that the basis of the phenotypic differences lies in the realm of gene regulation (King and Wilson 1975).

Direct evidence in support of the regulatory hypothesis has recently been provided by a number of comparative microarray studies showing that significant differences in gene expression patterns exist between humans and chimpanzees, especially in organs (e.g., brain and testes) and functions (e.g., cognitive ability and fertility) directly related to some of the major phenotypic traits distinguishing the two species (Li et al. 2001; Khaitovich et al. 2005). Questions remain, however, concerning the genetic basis of the

differences in gene regulation that separates humans from chimpanzees. One hypothesis is that the substantial INDEL variation that exists between humans and chimpanzees may contribute significantly to the regulatory differences between the species (Britten 2002; Polavarapu et al. 2006). To test this hypothesis, we categorized the INDEL variation existing between humans and chimpanzees that is located in or around genes and determined if this variation is significantly linked with species differences in gene expression. Our results indicate that such an association does exist and that it is attributable to both interspersed (predominately retrotransposon) and non-interspersed associated INDEL variation. The majority of this INDEL variation is attributed to lineage specific insertions and predominantly to insertions within the human lineage (Table 4.2 and Table 4.6).

We found that relatively little of the human-chimpanzee INDEL variation maps to the upstream region of genes. Since many essential control sequences are known to map upstream of the transcriptional start site of eukaryotic genes (Levine and Tjian 2003), most of these cis-regulatory sequences likely evolved prior to the divergence of chimpanzees and humans from a common ancestor ~ 6 million years ago. Thus, a possible explanation for the paucity of upstream INDEL variation between chimpanzees and humans in the upstream regions of genes is that most INDELs arising in this region are non-adaptive and thus rapidly removed by natural selection.

We found that the majority of the INDEL variation between humans and chimpanzees that is significantly correlated with differences in gene expression maps to introns. While

some of the variation in gene expression associated with human intronic insertions may be due to the incorporation of these inserted sequences into mRNAs (affecting RNA stability, etc.), the majority are most likely exerting enhancer or other cis-regulatory effects. Several examples of such insert-mediated cis-regulatory mutations have been previously documented in mice, rats and humans (Rothenburg et al. 2002; Yamada et al. 2006; Illarionova et al. 2007).

Our finding is that most of the INDEL variation between humans and chimpanzees which is associated with differential gene expression is attributable to human insertions. This finding is interesting for two reasons. First, it is consistent with the view that much of the divergence in gene expression that exists between chimpanzees and humans may have been driven by accelerated regulatory evolution within the human lineage (Enard et al. 2002; Gu and Gu 2003; Prabhakar et al. 2006; Wang et al. 2007). Second, it suggests that at least with respect to the evolutionary contribution of INDELs to chimpanzee-human divergence in gene expression, selection operating on *de novo* mutations may have been more important than selection operating on standing variation pre-existing in common ancestral populations. While previous analyses of gene expression and protein coding sequence variation between chimpanzees and humans revealed a pattern consistent with neutral evolution and negative selection (Khaitovich et al. 2005), our findings suggest that INDELs may have been a positive driving force behind human regulatory evolution.

CONCLUSIONS

In this study we tested to see whether there is a significant association between INDEL variation and differential gene expression between humans and chimpanzees. Our results indicate that indeed such an association does exist, and this association is more predominant with INDELs present in the intronic region of the genes. INDELs containing both interspersed (predominantly) and non-interspersed sequences contribute to differential expression, and we conclude that differences between the two species may be due to insertions of these sequences in the human lineage.

MATERIALS AND METHODS

Initial datasets

The database available at the UCSC Genome bioinformatics web site (Kent et al. 2002) was used to generate CGs and HGs datasets. These datasets contain genomic coordinates for CGs and HGs of sizes ranging from 80bp to 12,000bp. The CG dataset was generated by aligning the chimpanzee genome against the human genome build hg16 (July 2003) (Karolchik et al. 2003; Karolchik et al. 2004), and the HG dataset by aligning the human genome against the chimpanzee genome build panTro1 (Nov 2003). The CG and HG dataset coordinates were updated to the hg18 (Mar 2006) and panTro2 (Mar 2006) versions of the human and chimpanzee genome respectively, using genome browser utilities, Batch Coordinate Conversion liftOver tool (http://genome.ucsc.edu/util.html). Genomic sequences corresponding to the updated gap coordinates were downloaded from the UCSC genome database. Some of the gap sequences (76 CGs and 2581 HGs) not represented in the new versions of genome assemblies were removed in this process.

Identification of INDELs

Human genome (hg18), Chimpanzee genome (panTro2) and Macaque genome (rheMac2) sequences were downloaded from the UCSC genome database ftp website (ftp://hgdownload.cse.ucsc.edu/goldenPath). These genomes were aligned (see below), and in-house Perl scripts were used to process the alignment output files.

CG (HG) coordinates were searched in the multiple (pair-wise and three-way) alignments against the human (chimpanzee) and macaque genomes using in-house Perl scripts. Characterization of a gap as a chimpanzee (human) deletion or human (chimpanzee) insertion was determined using macaque as out-group (Figure 4.1).

(**Note:** HGs and CGs characterized as partial deletions or partial insertions due to incomplete sequencing of the macaque (out-group) genome were removed from the analysis.)

Characterization of sequences contained in INDELs

The Repeat Masker program (http://www.repeatmasker.org) was used to identify all interspersed repeats in the INDEL sequences. These were further classified according to the type of interspersed repeats, i.e. SINE, LINE, ERV, SVA or DNA elements. INDEL sequences consisting of more than one type of interspersed repeats (e.g. ER inserted within a LINE element) were classified as Mosaic elements (ME). The tandem repeat finder program (Benson 1999) was used to identify tandem repeat sequences within the INDELs characterized as non-interspersed sequences (INDELs not containing

interspersed repeat sequences). The remainder of the non-interspersed sequences was classified as unique sequences.

Association of human and chimpanzee genes with INDEL variation

The genomic coordinates for the genic regions (transcription start, transcription end, coding sequence (CDS) start, CDS end, intron start, intron end, exon start and exon end) for the human and chimpanzee Ensembl genes were downloaded from the UCSC Genome Bioinformatics web site (Kent et al. 2002). For each gene, 5000 bps upstream region of the transcription start site and 5000 bps downstream of the transcription end site were included in the genic region to correct for the comparatively longer average length of introns (Hong et al. 2006). An INDEL was considered to be associated with the gene if the genomic coordinates of the INDEL overlapped with the coordinates of the genic region. In-house Perl scripts were used to match these coordinates.

Analysis of microarray gene expression data

The human-chimpanzee gene expression raw data from five different tissues (i.e. brain, heart, liver, kidney and testis) in 6 humans and 5 chimpanzees were obtained from a previous microarray (Affymetrix HGU1332plus arrays) study (Khaitovich et al. 2005). The expression data was analyzed using MAS 5.0 normalization method encoded in the Affymetrix function library of the Bioconductor package (www.bioconductor.org) developed for R statistical programming environment (www.rproject.org) Genes with significant differences in affy probe sequences between humans and chimpanzees and with inconsistent hybridization patterns within samples in a species were excluded from

the analysis. Probe-sets having at least one presence call in any of the samples, either from human or chimpanzee, were called as detected and used for further analysis.

The expression values of these genes were normalized across samples by Z-score calculation using Spotfire Decision Site software (www.spotfire.com). Analysis of variance (ANOVA) was used to identify genes whose expression was significantly different ($p<0.05$) between human and chimpanzees in each of the tissues.

Categories of genes associated with INDEL variation between humans and chimpanzees

Genes associated with INDELs were grouped in two ways. First, they were grouped with respect to the type of sequences contained in the INDEL: a) interspersed repeat associated INDELs (>99% being retrotransposon sequences) or b) non–interspersed repeat associated INDELs. Second, genes associated with INDELs were grouped with respect to the location of the INDEL: intron, exon, upstream (within 5000bp upstream of transcription start site), or downstream (within 5000bp downstream of transcription termination site) genic locations. Some genes were grouped into more than one category (e.g., an INDEL in an upstream region that extends into the first intron). Genes located around INDELs comprised of retrotransposon sequences were further divided according to the type of retrotransposons (SINEs, LINEs, ERVs, SVAs or MEs). Genes associated with more than one class of INDELs were counted multiple times.

Correlating INDEL variation with differential gene expression

Differences in gene expression between chimpanzee and humans in each of the five tissues were partitioned for differentially expressed genes associated with INDELs vs. non-differentially expressed genes associated with INDELs. Proportions tests were used to determine the significance of the association of INDEL variation with differences in gene expression ($p < 0.05$) between the two species. The association was confirmed with a chi-square test of independence.

LIST OF ABBREVIATIONS:

INDEL: insertion/ deletion; CIs: chimpanzee insertions; CDs: chimpanzee deletions; HIs: human insertions; HDs: human deletions; HGs: human gaps; CGs: chimpanzee gaps; REs: retrotransposon sequences; NISs: non-interspersed sequences; SINEs: short interspersed nuclear elements; LINEs: long interspersed nuclear elements; ERVs: endogenous retroviruses; SVAs: biologically active composite elements consisting of fragments of SINE, VNTRs-variable number of tandem repeats, and Alu elements; ME: Mosaic elements; CERV/PTERV: chimpanzee endogenous viruses; MYA: million years ago; CDS: coding sequences; hg: human genome; panTro: chimpanzee genome; rheMac2: macaque genome; ANOVA: Analysis of variance; DE: differentially expressed genes; Exp: genes not differentially expressed.

CHAPTER 5

CONCLUSION

Human and chimpanzees differ in a number of phenotypic traits, chief among them are an increased size of the human brain (Carroll 2005; de Sousa and Wood 2007) and an increased propensity to cancer in humans (Seibold 1973). Apoptosis is known to play a role during neuronal development (Kuida et al. 1996; Honarpour et al. 2001; Joza et al. 2001), as well as disease progression to cancer (Kerr et al. 1972). The research in this study looks at differences in apoptotic function between humans and chimpanzees, to understand the basis of these phenotypic differences. In addition, this study also addresses the genetic basis of the phenotypic differences between the two species by looking at the Insertion/Deletion (INDEL) sequence variation that exists between the two species.

In this study, we saw that a number of genes were differentially expressed between humans and chimpanzees in five different tissues (brain, testis, heart, liver and kidney). To directly explore the consequences of the differences in gene expression, the gene expression differences between humans and chimpanzee brains were overlaid on known pathways. Results from the pathway analysis, suggested that the humans have a reduced apoptotic function in the brain, relative to chimpanzees. The reduced apoptotic function in the human brain, relative to chimpanzees, may have partly resulted in an increased size of the human brain.

The reduced apoptotic function in the human brain may have other consequences, besides it's potential contribution to brain size. A reduction in apoptotic function may serve to enhance longevity in humans. In the brain, all the neurons are produced during the early developmental years (Oomen et al. 2009). Newer neurons are not formed later in life. For a very long-lived species like *Homo sapiens* (Finch 2010), a reduced apoptotic function may set the threshold for neurons relatively high compared to chimpanzees. This high threshold may ensure that humans don't lose neurons that cannot be replaced later in life.

When the gene expression analysis was extended to other tissues (e.g. testis, heart, liver and kidney), a similar trend of reduced apoptotic function in humans compared to chimpanzees was seen. These results suggested that the apoptotic function maybe generally reduced in humans. The reduced apoptotic function in humans may account for an increased propensity for cancer in humans compared to chimpanzees, as during cancer the apoptotic function is overall reduced (Hanahan and Weinberg 2000). The onset of cancer in humans typically takes place well after reproductive age (http://seer.cancer.gov/csr/1975_2005/).

This antagonistic role of a reduced apoptotic function in humans during different life stages is analogous to a model of antagonistic pleiotropy, which was proposed by George C. Williams in 1957 to explain the evolutionary model of ageing (Williams 1957). According to this model, alleles that increase the fitness of an organism during early formative years are selected for, however the same alleles may have negative effects later in life.

Based on the above results, we hypothesized that the selection for increased cognitive ability in humans (by means of a reduced apoptotic function) may have had, as a side effect, an increased propensity for cancer in humans (Arora et al. 2009). There are two lines of evidence that support this hypothesis. First, chimpanzees have lower incidences of cancer relative to humans (McClure et al. 1973). In humans, death caused by a number of neoplasms is ~20%, compared to the incidences of these cancers in chimpanzees, which is ~2% (Seibold et al. 1973). Second, an increased apoptotic function associated with Huntington's' and Parkinson's disease patients (Kiechle T et al. 2002) has been correlated with a decreased propensity for cancer in these patients (Eskenazi et al. 2007).

In order to test the hypothesis of a reduced apoptotic function in humans, a series of apoptotic functional assays were conducted on primary fibroblast cell lines from humans, chimpanzees and macaques. Overall, human cells had higher cell viability, decreased caspase-3/7 activity and reduced ability to initiate apoptosis, compared to chimpanzees and macaques. These results are consistent with a reduced apoptotic function model in humans relative to chimpanzees and macaques and suggest that the reduction in apoptosis in humans is an evolutionary derived condition within the human lineage, subsequent to the divergence of humans and chimpanzees from a common ancestor

An ideal way to further test the hypothesis of a reduced apoptotic function in humans would be to directly carry out tests on tissues obtained from human and chimpanzee samples. It would be even more interesting to compare samples obtained from human and chimpanzee cancer samples. However, accessibility to such tissue is extremely difficult,

given that there are low rates of cancer in chimpanzees and that chimpanzees are an endangered species.

During brain development (Kiecker and Lumsden 2005) as well as cancer progression (Hanahan and Weinberg 2000), a number of pathways, other than apoptosis, are involved. By looking at pathways like Notch and Wnt signaling, it could be determined whether, just like apoptosis, these pathways show differences in their function between humans and chimpanzees. These differences in function could then be used to further confirm the basis of an increased brain size in humans, which is accompanied by an increased propensity for cancer in humans.

To understand the genetic basis of the phenotypic differences between humans and chimpanzees, the INDEL variation between the two species was analyzed and correlated to differences in gene expression. Human and chimpanzees are ~98.5% identical at the nucleotide level (95% if INDEL variation is considered between the two species) (Mikkelsen et al. 2005). In spite of the similarity between the two species at the nucleotide level, there are a number of phenotypic differences between the two species. A number of studies have suggested that the genetic basis of the differences between humans and chimpanzees lies at the level of the INDEL variation, which may contribute to the regulatory evolution that exists between the two species (Britten 2002; Frazer et al. 2003; Polavarapu et al. 2006).

In the INDEL analysis study, we saw a significant association of differential gene expression between the humans and chimpanzees with INDEL variation in various tissues. A majority of the INDEL variation was due to human-specific insertions and was mapped to the intronic region of the genes. Most of the INDELs were composed of either interspersed sequences (predominantly retrotransposons) or non-interspersed sequences, both of which are known to contribute to differences in gene regulation (McDonald 1993; Hammock and Young 2005; Mills et al. 2006a). These results are consistent with the findings that INDELs are significant contributors of variation between the two species (Volfovsky et al. 2009) and that INDELs composed of both interspersed (Britten 1997; Polavarapu et al. 2006) and non-interspersed sequences (Hammock and Young 2005) contribute to this variation.

The INDEL analysis has systematically correlated INDEL variation with gene expression differences between humans and chimpanzees. This study can be used a model to understand the role that INDEL variation may play in causing differential expression genes in diseases like cancer.

APPENDIX A

SUPPLEMENTARY INFORMATION FOR CHAPTER 2

Table A.1: Genes differentially expressed between human and chimpanzee brains associated, with the Huntington's and Parkinson's signaling pathways. The log ratio values of gene expression difference in human relative to chimpanzee are shown. A negative value indicates down-regulation in humans relative to chimpanzees while a positive value indicates up-regulation in humans relative to chimpanzee.

Gene Symbol	Description	Location in Human Genome	Pathway	Log ratio
AKT3	v-akt murine thymoma viral oncogene homolog 3 (protein kinase B, gamma)	chr1q43-q44	Huntington's	-0.34
AP2A2	adaptor-related protein complex 2, alpha 2 subunit	chr11p15.5	Huntington's	0.37
ATF2	activating transcription factor 2	chr2q32	Huntington's	0.22
BDNF	brain derived neurotrophic factor	chr11p13	Huntington's	-0.89
BET1L	blocked early in transport 1 homolog (*S. cerevisiae*)-like	chr11p15.5	Huntington's	-1.18
CACNA1B	Calcium channel voltage-dependent, type L, alpha 1B subunit	chr9q34	Huntington's	0.66
CAPN2	calpain 2	chr1q41-q42	Huntington's	-0.63
CAPN3	calpain 3	chr15q15.1-q21.1	Huntington's	1.22
CAPN5	calpain 5	chr11q14	Huntington's	-0.51
CAPN 7	calpain 7	chr3p24	Huntington's	-0.39
CASP2	caspase 2	chr7q34-q35	Huntington's	-1.44
CASP6	caspase 6	chr4q25	Huntington's	-1.78
CASP9	caspase 9	chr1p36.3-p36.1	Huntington's/ Parkinson's	-0.35
CLTC	clathrin, heavy chain (Hc)	chr16q22.1	Huntington's	-0.21
CPLX2	complexin 2	chr5q35.2	Huntington's	-1.26
CREB1	cAMP responsive element binding protein 1	chr2q34	Huntington's	0.58
CREBBP	CREB binding protein	chr16p13.3	Huntington's	-2.01
CYCS	cytochrome C, somatic	chr7p15.3	Huntington's /Parkinson's	-1.18

Table A.1 continued

DNAJB14	DnaJ (Hsp40) subfamily B member 14	chr4q23	Huntington's	0.55
DNM1	Dynamin 1	chr9q34	Huntington's	0.43
DYNC1I2	dynein, cytoplasmic 1intermediate chain 2	chr2q31.1	Huntington's	0.52
EGFR	epidermal growth factor receptor 1	chr7p12	Huntington's	0.87
FRAP1	FK506 binding protein 12-rapamycin associated protein 1	chr1q36.2	Huntington's	-0.61
GNAQ	guanine nucleotide binding protein (G protein), q polypeptide	chr9q21	Huntington's	0.67
GNB1	guanine nucleotide binding protein (B protein), beta polypeptide 1	chr1p36.33	Huntington's	-0.38
GNB4	guanine nucleotide binding protein (G protein), beta polypeptide 4	chr3q26.2-q26.33	Huntington's	0.84
GNB2L1	guanine nucleotide binding protein (G protein), beta polypeptide 2- like 1	chr5q35.3	Huntington's	0.59
GNG4	guanine nucleotide binding protein (G protein), gamma 4	chr1q42.3	Huntington's	1.15
GNG7	guanine nucleotide binding protein (G protein),gamma 7	chr19p13.3	Huntington's	0.98
GNG12	guanine nucleotide binding protein (G protein),gamma 12	chr1p31.3	Huntington's	1.94
GOSR1	golgi SNAP receptor complex member 1	chr17q11	Huntington's	1.45
GRB2	growth factor receptor-bound protein 2	chr17q24-q25	Huntington's	-0.46
GRIN2B	glutamate receptor, ionotrophic , N-Methyl D-aspartate 2B	chr12p12	Huntington's	-1.91
HDAC6	histone deacetylase 6	chrX11.23	Huntington's	-0.52
HDAC11	histone deacetylase 11	chr3p25.1	Huntington's	0.45
HDAC2	histone deacetylase 2	chr6q21	Huntington's	1.31
HIP1	Huntington interacting protein 1	chr7q11.23	Huntington's	-0.49

Table A.1 continued

HSPA2	heat shock 70kDa protein 2	chr14q24.1	Huntington's	1.91
HSPA5	heat shock 70kDa protein 5	chr9q33-q34.1	Huntington's	3.54
HSPA9	heat shock 70kDa protein 9	chr5q31.1	Huntington's	4.51
IFT57	intraflagellar transport 57 homolog	chr3q13.12	Huntington's	-0.38
ITPR1	inositol 1,4,5-triphosphate receptor type 1	chr3p26-p25	Huntington's	1.67
JUN	Jun oncogene	chr1p32-p31	Huntington's	-0.46
MAP2K7	Mitogen activated protein kinase kinase 7	chr19p13.3-p13.2	Huntington's	-0.47
MAPK1	Mitogen activated protein kinase 1	chr22q11.2|22q11.21	Huntington's	-0.57
MAPK3	Mitogen activated protein kinase 1	chr16p11.2	Huntington's/ Parkinson's	0.68
MAPK8	Mitogen activated protein kinase 8	chr10q11.2	Huntington's	-0.51
MAPK9	Mitogen activated kinase 9	chr5q35	Huntington's	1.03
MAPK10	Mitogen activated kinase 10	chr4q22.1-q22.3	Huntington's	-0.33
NAPA	N-ethylmaleimide-sensitive factor attachment protein, alpha	chr19q13.32	Huntington's	-0.99
NAPG	N-ethylmaleimide-sensitive factor attachment protein, gamma	chr18p11.22	Huntington's	1.32
NCOR1	Nuclease receptor corepressor 1	chr17p11.2	Huntington's	-0.42
NCOR2	Nuclease receptor corepressor 2	chr12q24	Huntington's	-0.44
PIK3C3	phosphoinositide-3-kinase, class 3	chr18q12.3	Huntington's	1.82
PIK3C2A	phosphoinositide-3-kinase, class 3, catalytic alpha polypeptide	chr11p15.5-p14	Huntington's	1.19
PIK3CA	Phosphoinositide-3-kinase, catalytic, alpha polypeptide	chr3q26.3	Huntington's	-0.74
PIK3CB	phosphoinositide-3-kinase, catalytic, beta polypeptide	chr3q22.3	Huntington's	0.55
PIK3CD	phosphoinositide-3-kinase, catalytic, delta polypeptide	chr1p36.2	Huntington's	0.72
PIK3R1	phosphoinositide-3-kinase, regulatory subunit 1 (p85 alpha)	chr5q13.1	Huntington's	-1.10
PLCB1	phospholipase C, beta 1 (phosphoinositide-specific)	chr20p12	Huntington's	-1.88

Table A.1 continued

PLCB4	phospholipase C, beta 4	chr20p12	Huntington's	-0.75
POLR2A	polymerase (RNA) II (DNA directed) polypeptide A, 220kDa	chr17p13.1	Huntington's	0.55
POLR2G	polymerase (RNA) II (DNA directed) polypeptide G	chr11q13.1	Huntington's	0.34
POLR2H	polymerase (RNA) II (DNA directed) polypeptide H	chr3q28	Huntington's	-0.21
POLR2I	polymerase (RNA) II (DNA directed) polypeptide I, 14.5kDa	chr19q12	Huntington's	0.94
POLR2J	polymerase (RNA) II (DNA directed) polypeptide J, 13.3kDa	chr7q22.1	Huntington's	1.4
POLR2K	polymerase (RNA) II (DNA directed) polypeptide K, 7.0kDa	chr8q22.2	Huntington's	1.04
PRKCE	protein kinase C, epsilon	chr2p21	Huntington's	-0.74
PRKCI	protein kinase C, iota	chr3q26.3	Huntington's	1.2
PRKCQ	protein kinase C, theta	chr10p15	Huntington's	0.55
RASA1	RAS p21 protein activator (GTPase activating protein) 1	chr5q13.3	Huntington's	0.25
RCOR1	REST corepressor 1	chr14q32.32	Huntington's	-0.30
SDHB	succinate dehydrogenase complex, subunit B, iron sulfur (Ip)	chr1p36.1-p35	Huntington's	-1.43
SH3GL3	SH3-domain GRB2-like 3	chr15q24	Huntington's	0.48
SHC1	SHC (Src homology 2 domain containing) transforming protein 1	chr1q21	Huntington's	-0.21
SNAP25	synaptosomal-associated protein, 25kDa	chr20p12-p11.2	Huntington's	-0.31
TAF4	TAF4 RNA polymerase II, TATA box binding protein (TBP)-associated factor, 135kDa	chr20q13.33	Huntington's	-1.14
TAF9B	TAF9B RNA polymerase II, TATA box binding protein (TBP)-associated factor, 31kDa	chrXq13.1-q21.1	Huntington's	-0.75
TBP	TATA box binding protein	chr6q27	Huntington's	-0.39
UBB	ubiquitin B	chr17p12-11.2	Huntington's	0.27

Table A.1 continued

UBC	ubiquitin C	chr12q24.3	Huntington's	-0.31
ZDHHC17	Zinc finger, DHCC-type containing 17	chr12q21.2	Huntington's	-0.41
DIABLO	Diablo homolog (drosophila)	chr12q24.31	Huntington's /Parkinson's	-0.51
MAPK12	mitogen activated protein kinase 12	chr22q13.33	Huntington's /Parkinson's	-1.04
SEPT5	septin 5	chr22q11.21	Parkinson's	0.43
SNCA	synuclein, alpha	chr4q21	Huntington's /Parkinson's	1.47
UCHL1	Ubiquitin carboxyl terminal esterase L1	chr4p14	Parkinson's	-0.59

Genetic basis of the observed differences in expression of apoptotic genes between humans and chimpanzees

In a preliminary effort to discern the genetic basis of the observed differences in the expression of apoptotic genes, a comparative sequence alignment was performed between the 37 apoptotic genes differentially expressed between human and chimpanzee using the macaque as an out-group. Remarkably, none of the 37 differentially expressed genes displayed variation within their coding regions. In contrast, significant Insertion/Deletion (INDEL) variation was detected between the two species in introns and upstream regions at 34 of the 37 genes examined (Table A.2, Figure A.1) as analyzed on the UCSC genome browser (Kent et al. 2002). A number of these genes with INDEL variation, e.g., CYCS, AIFM1, DIABLO, CASP6, CASP9, BID, BIRC4 and HTRA2, are involved in the intrinsic apoptotic pathway, and a majority of these have this INDEL variation in the upstream region. Two of these genes, e.g. DIABLO and HTRA2, have transcription factor binding sites (TFBS) missing in the chimpanzee sequences.

Table A.2: Apoptotic pathway genes differentially expressed between the human and chimpanzee brains are associated with INDEL variation.
No nucleotide or INDEL variation was detected within the coding regions (exons) of the 37 apoptotic pathway genes differentially expressed between the human and chimpanzee brain. In contrast, substantial INDEL variation was detected in introns and/or upstream regions (1-5000 bp 5' to the transcriptional start site) in 34 of the 37 genes. Pair wise alignments of mRNA sequences (Needleman and Wunsch 1970) were used to search for nucleotide variation. INDEL variation was identified using the macaque as an out-group. (H = human, C = chimpanzee, ↓ = Insertion, ↑ = Deletion, * indicates possible underestimate of INDEL variation in upstream region due to poor sequence quality). Size of INDEL and distinguishing feature is given in parentheses: TFBS= Transcription factor binding sites; SINE = short interspersed nuclear element; LINE = long interspersed nuclear element; SVA= composite of SINE, Variable number tandem repeat and Alu; LTR = long terminal repeat retrotransposon element; CpG= Cytosine phosphate guanine region.

Gene Symbol	INDEL VARIATION	
	INTRON	UPSTREAM REGION
1.AIFM1*	None	1 C ↑(3328 bps long)
2.BID*	1 H ↑ (184 bps; tandem repeat)	None
3.BIRC4*	1 C ↑ (330 bps; SINE element)	1 C ↑ (1240 bps; 3 SINE elements)
4.CAPN2	None	1 H ↑ (10 bps; part of SINE element)
5.CAPN3	None	1 C ↑ (401 bps; 2 SINE elements)
6.CAPN5	1 H ↓ (2337 bps; Other repeat)	None
7.CAPN7	1 C ↓ (82 bps; simple repeat)	None
8.CASP2	1 C ↑ (21 bps; part of LTR element)	None
9.CASP6	None	1 C ↑ (1630 bps; 3 SINE elements)
10.CASP9	None	1 C ↑ (10 bps; part of SINE element)
11.CHUK	None	None
12. CYCS	None	1 H↑ (4711 bps; 4 SINE elements)

Table A.2 continued

13. DDR1	None	1 H↑(23 bps; part of LINE element)
14.DDR2	2 H ↓ (135 bps; simple repeat and 2209 bps; LINE element)	None
15.DFFA	None	None
16.DIABLO*	None	1 C ↑ (840 bps; part CpG region along with 3 TFBS missing)
17.EGFR	1 C ↓ (313 bps; SINE element)	None
18.FGFR2	1 H ↓ (295 bps; SINE element)	1 H ↓ (92 bps; SINE element)
19.HTRA2*	None	1 C ↑ (854 bps; part CpG region along with 4 TFBS missing)
20.IKBKB	1 H ↑ (313 bps; SINE element)	None
21.KRAS	1 C ↑ (216 bps; part SINE element missing)	None
22. MAP2K2	1 H ↓ (90 bps; simple repeat)	1 C ↑ (339 bps; part CpG region along with 1 TFBS missing)
23.MAP2K7*	None	3 C ↑ (311 and 123 bps; 2 SINE elements and 932 bps; part of CpG region along with 2 TFBS missing)
24.MAP3K5	1 C ↓(134 bps; simple repeat)	1 H ↑ (295 bps; 1 SINE element)
25.MAP4K4	2 H ↓(315 bps; 1 SINE element and 144 bps; simple repeat)	None
26.MAPK1	1 H ↓(94 bps; part of SINE element)	1 H ↑ (171 bps; part of a SINE element)
27.MAPK3	None	1 H ↑ (1959 bps; 5 SINE elements)
28.MAPK8	None	1 C ↑(36bps)
29.MRAS	None	1 C ↑ (28 bps; part of a SINE element)
30.PARP1	None	1 C ↑ (20 bps; part of a SINE element)

Table A.2 continued

31.PDGFRA	1 H ↑ (1889 bps; SVA element)	None
32. PRKCE	1 H ↑(112 bps; Simple repeat)	None
	5 H ↓ (314 bps, 318 bps, 314 bps, 317 bps and 308 bps; 5 SINE elements)	
33.RAF1	None	1 H ↑(26 bps; part of SINE element)
34.RELA	None	1 C ↑ (47 bps; part of SINE element)
35.SPTAN1	1 C ↑ (299 bps; SINE element missing)	None
36.TEK	None	None
37.TYRO3	None	2 C ↑ (1444 bps, 776 bps; part of SINE element missing in one ↑)

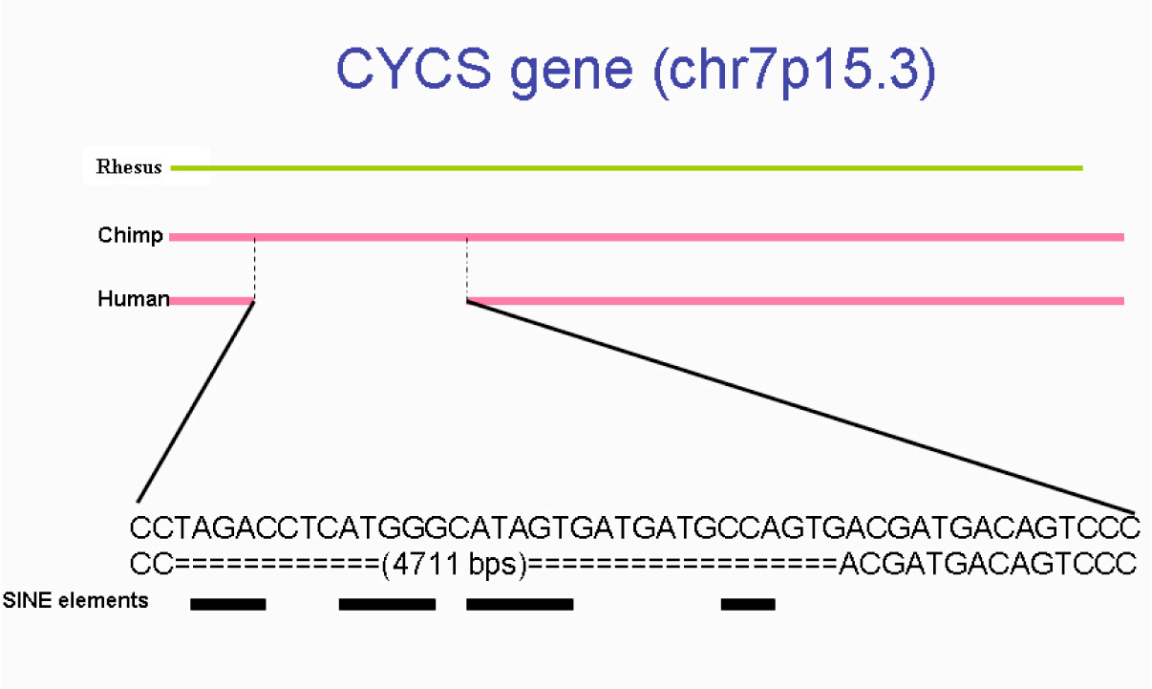


Figure A.1: An example of INDEL variation located upstream of an apoptotic pathway gene e.g. cytochrome C (*CYCS*).
All alignments between human, chimpanzee and macaque (out-group) sequences were performed employing the Genome Browser Gateway [(Kent et al. 2002),Human, Chimpanzee and Macaque 2006 Assembly] and using the Needleman-Wunsch global alignment algorithm (Needleman and Wunsch 1970). A 4.7 kb region containing four SINE elements has been deleted from a region upstream of the human cytochrome C (*CYCS*) gene since the divergence of humans and chimpanzees from a common ancestor. Significant differences in *CYCS* expression were detected in human and chimpanzee brains despite the fact that no nucleotide differences exist between the two species within the coding region (exons) of the gene.

APPENDIX B

SUPPLEMENTARY INFORMATION FOR CHAPTER 4

Table B.1: Number of genes associated with INDELs and non-INDELs. The numbers in the four right-most columns are the number of genes differentially expressed (DE) or not (Exp) in each tissue that are associated with INDELs and non-INDELs

	INDELs		non-INDELs	
Tissue (Total number of genes detected in each tissue)	**Number of DE genes**	**Number of Exp genes**	**Number of DE genes**	**Number of Exp genes**
Brain (14133)	2266	2153	4618	5096
Testis (15445)	3438	1256	7365	3386
Heart (13497)	2233	1948	4610	4706
Liver (13684)	1696	2466	3612	5910
Kidney (14059)	2179	2144	4410	5326

REFERENCES

Archidiacono N, Antonacci R, Marzella R, Finelli P, Lonoce A, Rocchi M. 1995. Comparative mapping of human alphoid sequences in great apes using fluorescence in situ hybridization. *Genomics* **25**(2): 477-484.

Arora G, Polavarapu N, McDonald JF. 2009. Did natural selection for increased cognitive ability in humans lead to an elevated risk of cancer? *Med Hypotheses* **73**(3): 453-456.

Ashkenazi A, Dixit VM. 1998. Death receptors: signaling and modulation. *Science* **281**(5381): 1305-1308.

Beniashvili DS. 1989. An overview of the world literature on spontaneous tumors in nonhuman primates. *J Med Primatol* **18**(6): 423-437.

Benson G. 1999. Tandem repeats finder: a program to analyze DNA sequences. *Nucleic Acids Res* **27**(2): 573-580.

Biro D, Inoue-Nakamura N, Tonooka R, Yamakoshi G, Sousa C, Matsuzawa T. 2003. Cultural innovation and transmission of tool use in wild chimpanzees: evidence from field experiments. *Anim Cogn* **6**(4): 213-223.

Blurton Jones NG, Hawkes K, O'Connell JF. 2002. Antiquity of postreproductive life: are there modern impacts on hunter-gatherer postreproductive life spans? *Am J Hum Biol* **14**(2): 184-205.

Bodamer MD, Gardner RA. 2002. How cross-fostered chimpanzees (Pan troglodytes) initiate and maintain conversations. *J Comp Psychol* **116**(1): 12-26.

Bonner TI, Birkenmeier EH, Gonda MA, Mark GE, Searfoss GH, Todaro GJ. 1982. Molecular cloning of a family of retroviral sequences found in chimpanzee but not human DNA. *J Virol* **43**(3): 914-924.

Bowen NJ, Jordan IK. 2002. Transposable elements and the evolution of eukaryotic complexity. *Curr Issues Mol Biol* **4**(3): 65-76.

Britten RJ. 1997. Mobile elements inserted in the distant past have taken on important functions. *Gene* **205**(1-2):177-82.

Britten RJ. 2002. Divergence between samples of chimpanzee and human DNA sequences is 5%, counting indels. *Proc Natl Acad Sci U S A* **99**(21): 13633-13635.

Brunet M, Guy F, Pilbeam D, Mackaye HT, Likius A, Ahounta D, Beauvilain A, Blondel C, Bocherens H, Boisserie JR et al. 2002. A new hominid from the Upper Miocene of Chad, Central Africa. *Nature* **418**(6894): 145-151.

Budd SL, Tenneti L, Lishnak T, Lipton SA. 2000. Mitochondrial and extramitochondrial apoptotic signaling pathways in cerebrocortical neurons. *Proc Natl Acad Sci U S A* **97**(11): 6161-6166.

Graves C.P. 1986. Systematics of great Apes. In *Comparative Primate Biology*, Vol 1 (ed. DRaE Swindler, J.), pp. 187-218. Alan R.Liss, New York.

Caccone A, Powell JR. 1989. DNA Divergence among Hominoids. *Evolution* **43**(5): 925-942.

Caceres M, Lachuer J, Zapala MA, Redmond JC, Kudo L, Geschwind DH, Lockhart DJ, Preuss TM, Barlow C. 2003. Elevated gene expression levels distinguish human from non-human primate brains. *Proc Natl Acad Sci U S A* **100**(22): 13030-13035.

Cadet JL, Jayanthi S, McCoy MT, Vawter M, Ladenheim B. 2001. Temporal profiling of methamphetamine-induced changes in gene expression in the mouse brain: evidence from cDNA array. *Synapse* **41**(1): 40-48.

Calarco JA, Xing Y, Caceres M, Calarco JP, Xiao X, Pan Q, Lee C, Preuss TM, Blencowe BJ. 2007. Global analysis of alternative splicing differences between humans and chimpanzees. *Genes Dev* **21**(22): 2963-2975.

Carroll SB. 2003. Genetics and the making of Homo sapiens. *Nature* **422**(6934): 849-857.

Carroll SB. 2005. Evolution at two levels: on genes and form. *PLoS Biol* **3**(7): e245.

Cecconi F, Alvarez-Bolado G, Meyer BI, Roth KA, Gruss P. 1998. Apaf1 (CED-4 homolog) regulates programmed cell death in mammalian development. *Cell* **94**(6): 727-737.

Chan WY, Lorke DE, Tiu SC, Yew DT. 2002. Proliferation and apoptosis in the developing human neocortex. *Anat Rec* **267**(4): 261-276.

Chen FC, Li WH. 2001. Genomic divergences between humans and other hominoids and the effective population size of the common ancestor of humans and chimpanzees. *Am J Hum Genet* **68**(2): 444-456.

Chen FC, Chen CJ, Chuang TJ. 2007a. INDELSCAN: a web server for comparative identification of species-specific and non-species-specific insertion/deletion events. *Nucleic Acids Res* **35**(Web Server issue): W633-638.

Chen FC, Chen CJ, Li WH, Chuang TJ. 2007b. Human-specific insertions and deletions inferred from mammalian genome sequences. *Genome Res* **17**(1): 16-22.

Chen M, Ona VO, Li M, Ferrante RJ, Fink KB, Zhu S, Bian J, Guo L, Farrell LA, Hersch SM et al. 2000. Minocycline inhibits caspase-1 and caspase-3 expression and delays mortality in a transgenic mouse model of Huntington disease. *Nat Med* **6**(7): 797-801.

Cheng Y, Prusoff WH. 1973. Relationship between the inhibition constant (K1) and the concentration of inhibitor which causes 50 per cent inhibition (I50) of an enzymatic reaction. *Biochem Pharmacol* **22**(23): 3099-3108.

Cheung EC, Slack RS. 2004. Emerging role for ERK as a key regulator of neuronal apoptosis. *Sci STKE* **2004**(251): PE45.

Chicheportiche Y, Bourdon PR, Xu H, Hsu YM, Scott H, Hession C, Garcia I, Browning JL. 1997. TWEAK, a new secreted ligand in the tumor necrosis factor family that weakly induces apoptosis. *J Biol Chem* **272**(51): 32401-32410.

Chinnaiyan AM. 1999. The apoptosome: heart and soul of the cell death machine. *Neoplasia* **1**(1): 5-15.

Chou HH, Takematsu H, Diaz S, Iber J, Nickerson E, Wright KL, Muchmore EA, Nelson DL, Warren ST, Varki A. 1998. A mutation in human CMP-sialic acid hydroxylase occurred after the Homo-Pan divergence. *Proc Natl Acad Sci U S A* **95**(20): 11751-11756.

Cory S, Adams JM. 2002. The Bcl2 family: regulators of the cellular life-or-death switch. *Nat Rev Cancer* **2**(9): 647-656.

Cowan WM, Fawcett JW, O'Leary DD, Stanfield BB. 1984. Regressive events in neurogenesis. *Science* **225**(4668): 1258-1265.

Crow TJ. 2002. Handedness, language lateralisation and anatomical asymmetry: relevance of protocadherin XY to hominid speciation and the aetiology of psychosis. Point of view. *Br J Psychiatry* **181**: 295-297.

Darwin C. 1871. *The descent of man, and selection in relation to Sex*. John Murray.

Deininger PL, Batzer MA. 1999. Alu repeats and human disease. *Mol Genet Metab* **67**(3): 183-193.

Deininger PL, Batzer MA. 2002. Mammalian retroelements. *Genome Res* **12**(10): 1455-1465.

Deng X, Ruvolo P, Carr B, May WS, Jr. 2000. Survival function of ERK1/2 as IL-3-activated, staurosporine-resistant Bcl2 kinases. *Proc Natl Acad Sci U S A* **97**(4): 1578-1583.

de Sousa A, Wood B. 2007. The hominin fossil record and the emergence of the modern human central nervous system. In *Evolution of Nervous Systems. Vol. 4: Primates* (ed. Kaas JH, PreussTM), pp 291-336. Oxford: Elsevier.

Doolittle RF, Wooding GL, Lin Y, Riley M. 1971. Hominoid evolution as judged by fibrinopeptide structures. *J Mol Evol* **1**(1): 74-83.

Driver JA, Logroscino G, Buring JE, Gaziano JM, Kurth T. 2007. A prospective cohort study of cancer incidence following the diagnosis of Parkinson's disease. *Cancer Epidemiol Biomarkers Prev* **16**(6): 1260-1265.

Du C, Fang M, Li Y, Li L, Wang X. 2000. Smac, a mitochondrial protein that promotes cytochrome c-dependent caspase activation by eliminating IAP inhibition. *Cell* **102**(1): 33-42.

Du J, Zhu Y, Shanmugam A, Kenter AL. 1997. Analysis of immunoglobulin Sgamma3 recombination breakpoints by PCR: implications for the mechanism of isotype switching. *Nucleic Acids Res* **25**(15): 3066-3073.

Ehrenberg B, Montana V, Wei MD, Wuskell JP, Loew LM. 1988. Membrane potential can be determined in individual cells from the nernstian distribution of cationic dyes. *Biophys J* **53**(5): 785-794.

Elmore S. 2007. Apoptosis: a review of programmed cell death. *Toxicol Pathol* **35**(4): 495-516.

Enard W, Khaitovich P, Klose J, Zollner S, Heissig F, Giavalisco P, Nieselt-Struwe K, Muchmore E, Varki A, Ravid R et al. 2002a. Intra- and interspecific variation in primate gene expression patterns. *Science* **296**(5566): 340-343.

Enard W, Przeworski M, Fisher SE, Lai CS, Wiebe V, Kitano T, Monaco AP, Paabo S. 2002b. Molecular evolution of FOXP2, a gene involved in speech and language. *Nature* **418**(6900): 869-872.

Eskenazi BR, Wilson-Rich NS, Starks PT. 2007. A Darwinian approach to Huntington's disease: subtle health benefits of a neurological disorder. *Med Hypotheses* **69**(6): 1183-1189.

Evans PD, Anderson JR, Vallender EJ, Choi SS, Lahn BT. 2004a. Reconstructing the evolutionary history of microcephalin, a gene controlling human brain size. *Hum Mol Genet* **13**(11): 1139-1145.

Evans PD, Anderson JR, Vallender EJ, Gilbert SL, Malcom CM, Dorus S, Lahn BT. 2004b. Adaptive evolution of ASPM, a major determinant of cerebral cortical size in humans. *Hum Mol Genet* **13**(5): 489-494.

Eyre-Walker A, Keightley PD. 1999. High genomic deleterious mutation rates in hominids. *Nature* **397**(6717): 344-347.

Ferreira IL, Nascimento MV, Ribeiro M, Almeida S, Cardoso SM, Grazina M, Pratas J, Santos MJ, Januario C, Oliveira CR et al. 2010. Mitochondrial-dependent apoptosis in Huntington's disease human cybrids. *Exp Neurol* **222**(2): 243-255.

Finch CE. 2010. Evolution of the human lifespan and diseases of aging: Roles of infection,inflammation and nutrition. *PNAs* **107** (supp 1): 1718-1724.

Finnegan DJ. 1992. Transposable elements. *Curr Opin Genet Dev* **2**(6): 861-867.

Frazer KA, Chen X, Hinds DA, Pant PV, Patil N, Cox DR. 2003. Genomic DNA insertions and deletions occur frequently between humans and nonhuman primates. *Genome Res* **13**(3): 341-346.

Fu YH, Pizzuti A, Fenwick RG, Jr., King J, Rajnarayan S, Dunne PW, Dubel J, Nasser GA, Ashizawa T, de Jong P et al. 1992. An unstable triplet repeat in a gene related to myotonic muscular dystrophy. *Science* **255**(5049): 1256-1258.

Fujiyama A, Watanabe H, Toyoda A, Taylor TD, Itoh T, Tsai SF, Park HS, Yaspo ML, Lehrach H, Chen Z et al. 2002a. Construction and analysis of a human-chimpanzee comparative clone map. *Science* **295**(5552): 131-134.

Gagneux P, Varki A. 2001. Genetic differences between humans and great apes. *Mol Phylogenet Evol* **18**(1): 2-13.

Gearing M, Rebeck GW, Hyman BT, Tigges J, Mirra SS. 1994. Neuropathology and apolipoprotein E profile of aged chimpanzees: implications for Alzheimer disease. *Proc Natl Acad Sci U S A* **91**(20): 9382-9386.

Gibbs RA Rogers J Katze MG Bumgarner R Weinstock GM Mardis ER Remington KA Strausberg RL Venter JC Wilson RK et al. 2007. Evolutionary and biomedical insights from the rhesus macaque genome. *Science* **316**(5822): 222-234.

Goodman M. 1996. Epilogue: a personal account of the origins of a new paradigm. *Mol Phylogenet Evol* **5**(1): 269-285.

Gu J, Gu X. 2003. Induced gene expression in human brain after the split from chimpanzee. *Trends Genet* **19**(2): 63-65.

Hacia JG. 2001. Genome of the apes. *Trends Genet* **17**(11): 637-645.

Hamada H, Seidman M, Howard BH, Gorman CM. 1984. Enhanced gene expression by the poly(dT-dG).poly(dC-dA) sequence. *Mol Cell Biol* **4**(12): 2622-2630.

Hammock EA, Young LJ. 2005. Microsatellite instability generates diversity in brain and sociobehavioral traits. *Science* **308**(5728): 1630-1634.

Hanahan D, Weinberg RA. 2000. The hallmarks of cancer. *Cell* **100**(1): 57-70.

Hasler J, Strub K. 2006. Alu elements as regulators of gene expression. *Nucleic Acids Res* **34**(19): 5491-5497.

Hawkes K, O'Connell JF, Jones NG, Alvarez H, Charnov EL. 1998. Grandmothering, menopause, and the evolution of human life histories. *Proc Natl Acad Sci U S A* **95**(3): 1336-1339.

Haydar TF, Kuan CY, Flavell RA, Rakic P. 1999. The role of cell death in regulating the size and shape of the mammalian forebrain. *Cereb Cortex* **9**(6): 621-626.

Hellmann I, Zollner S, Enard W, Ebersberger I, Nickel B, Paabo S. 2003. Selection on human genes as revealed by comparisons to chimpanzee cDNA. *Genome Res* **13**(5): 831-837.

Herrmann E, Call J, Hernandez-Lloreda MV, Hare B, Tomasello M. 2007. Humans have evolved specialized skills of social cognition: the cultural intelligence hypothesis. *Science* **317**(5843): 1360-1366.

Hickey DA. 1982. Selfish DNA: a sexually-transmitted nuclear parasite. *Genetics* **101**(3-4): 519-531.

Hill MM, Adrain C, Duriez PJ, Creagh EM, Martin SJ. 2004. Analysis of the composition, assembly kinetics and activity of native Apaf-1 apoptosomes. *EMBO J* **23**(10): 2134-2145.

Hockenbery D, Nunez G, Milliman C, Schreiber RD, Korsmeyer SJ. 1990. Bcl-2 is an inner mitochondrial membrane protein that blocks programmed cell death. *Nature* **348**(6299): 334-336.

Holmes SE, Dombroski BA, Krebs CM, Boehm CD, Kazazian HH, Jr. 1994. A new retrotransposable human L1 element from the LRE2 locus on chromosome 1q produces a chimaeric insertion. *Nat Genet* **7**(2): 143-148.

Honarpour N, Gilbert SL, Lahn BT, Wang X, Herz J. 2001. Apaf-1 deficiency and neural tube closure defects are found in fog mice. *Proc Natl Acad Sci U S A* **98**(17): 9683-9687.

Hong X, Scofield DG, Lynch M. 2006. Intron size, abundance, and distribution within untranslated regions of genes. *Mol Biol Evol* **23**(12): 2392-2404.

Hubbard TJ, Aken BL, Beal K, Ballester B, Caccamo M, Chen Y, Clarke L, Coates G, Cunningham F, Cutts T et al. 2007. Ensembl 2007. *Nucleic Acids Res* **35**(Database issue): D610-617.

Huda A, Jordan IK. 2009. Epigenetic regulation of Mammalian genomes by transposable elements. *Ann N Y Acad Sci* **1178**: 276-284.

Huda A, Marino-Ramirez L, Landsman D, Jordan IK. 2009. Repetitive DNA elements, nucleosome binding and human gene expression. *Gene* **436**(1-2): 12-22.

Huxley TH. 1863. *Evidence as to Man's place in Nature*. Williams and Norgate.

Igney FH, Krammer PH. 2002. Death and anti-death: tumour resistance to apoptosis. *Nat Rev Cancer* **2**(4): 277-288.

Illarionova AE, Vinogradova TV, Sverdlov ED. 2007. Only those genes of the KIAA1245 gene subfamily that contain HERV(K) LTRs in their introns are transcriptionally active. *Virology* **358**(1): 39-47.

Jacobson MD, Weil M, Raff MC. 1997. Programmed cell death in animal development. *Cell* **88**(3): 347-354.

Jayanthi S, Deng X, Bordelon M, McCoy MT, Cadet JL. 2001. Methamphetamine causes differential regulation of pro-death and anti-death Bcl-2 genes in the mouse neocortex. *FASEB J* **15**(10): 1745-1752.

Jones N, Dumont DJ. 2000. Tek/Tie2 signaling: new and old partners. *Cancer Metastasis Rev* **19**(1-2): 13-17.

Joza N, Susin SA, Daugas E, Stanford WL, Cho SK, Li CY, Sasaki T, Elia AJ, Cheng HY, Ravagnan L et al. 2001. Essential role of the mitochondrial apoptosis-inducing factor in programmed cell death. *Nature* **410**(6828): 549-554.

Kaessmann H, Wiebe V, Weiss G, Paabo S. 2001. Great ape DNA sequences reveal a reduced diversity and an expansion in humans. *Nat Genet* **27**(2): 155-156.

Kapitonov VV, Jurka J. 2003. Molecular paleontology of transposable elements in the Drosophila melanogaster genome. *Proc Natl Acad Sci U S A* **100**(11): 6569-6574.

Karaman MW, Houck ML, Chemnick LG, Nagpal S, Chawannakul D, Sudano D, Pike BL, Ho VV, Ryder OA, Hacia JG. 2003. Comparative analysis of gene-expression patterns in human and African great ape cultured fibroblasts. *Genome Res* **13**(7): 1619-1630.

Karlsson KA. 1995. Microbial recognition of target-cell glycoconjugates. *Curr Opin Struct Biol* **5**(5): 622-635.

Karolchik D, Baertsch R, Diekhans M, Furey TS, Hinrichs A, Lu YT, Roskin KM, Schwartz M, Sugnet CW, Thomas DJ et al. 2003. The UCSC Genome Browser Database. *Nucleic Acids Res* **31**(1): 51-54.

Karolchik D, Hinrichs AS, Furey TS, Roskin KM, Sugnet CW, Haussler D, Kent WJ. 2004. The UCSC Table Browser data retrieval tool. *Nucleic Acids Res* **32**(Database issue): D493-496.

Kataoka T, Schroter M, Hahne M, Schneider P, Irmler M, Thome M, Froelich CJ, Tschopp J. 1998. FLIP prevents apoptosis induced by death receptors but not by perforin/granzyme B, chemotherapeutic drugs, and gamma irradiation. *J Immunol* **161**(8): 3936-3942.

Kent WJ, Sugnet CW, Furey TS, Roskin KM, Pringle TH, Zahler AM, Haussler D. 2006. The human genome browser at UCSC. *Genome Res* **12(**6): 996-1006.

Kerr JF, Wyllie AH, Currie AR. 1972. Apoptosis: a basic biological phenomenon with wide-ranging implications in tissue kinetics. *Br J Cancer* **26**(4): 239-257.

Khaitovich P, Hellmann I, Enard W, Nowick K, Leinweber M, Franz H, Weiss G, Lachmann M, Paabo S. 2005. Parallel patterns of evolution in the genomes and transcriptomes of humans and chimpanzees. *Science* **309**(5742): 1850-1854.

Khaitovich P, Muetzel B, She X, Lachmann M, Hellmann I, Dietzsch J, Steigele S, Do HH, Weiss G, Enard W et al. 2004. Regional patterns of gene expression in human and chimpanzee brains. *Genome Res* **14**(8): 1462-1473.

Kidwell MG, Lisch DR. 2000. Transposable elements and host genome evolution. *Trends Ecol Evol* **15**(3): 95-99.

Kidwell MG, Lisch DR. 2001. Perspective: transposable elements, parasitic DNA, and genome evolution. *Evolution* **55**(1): 1-24.

Kiechle T, Dedeoglu A, Kubilus J, Kowall NW, Beal MF, Friedlander RM, Hersch SM, Ferrante RJ. 2002. Cytochrome C and caspase-9 expression in Huntington's disease. *Neuromolecular Med* **1**(3): 183-195.

Kiecker C, Lumsden A. 2005. Compartments and their boundaries in vertebrate brain development. *Nat Rev Neurosci* **6**(7): 553-564.

King MC, Wilson AC. 1975. Evolution at two levels in humans and chimpanzees. *Science* **188**(4184): 107-116.

Kischkel FC, Hellbardt S, Behrmann I, Germer M, Pawlita M, Krammer PH, Peter ME. 1995. Cytotoxicity-dependent APO-1 (Fas/CD95)-associated proteins form a

death-inducing signaling complex (DISC) with the receptor. *EMBO J* **14**(22): 5579-5588.

Klivenyi P, Ferrante RJ, Matthews RT, Bogdanov MB, Klein AM, Andreassen OA, Mueller G, Wermer M, Kaddurah-Daouk R, Beal MF. 1999. Neuroprotective effects of creatine in a transgenic animal model of amyotrophic lateral sclerosis. *Nat Med* **5**(3): 347-350.

Kouprina N, Pavlicek A, Mochida GH, Solomon G, Gersch W, Yoon YH, Collura R, Ruvolo M, Barrett JC, Woods CG et al. 2004. Accelerated evolution of the ASPM gene controlling brain size begins prior to human brain expansion. *PLoS Biol* **2**(5): E126.

Kroemer G, Reed JC. 2000. Mitochondrial control of cell death. *Nat Med* **6**(5): 513-519.

Kuan CY, Roth KA, Flavell RA, Rakic P. 2000. Mechanisms of programmed cell death in the developing brain. *Trends Neurosci* **23**(7): 291-297.

Kuan CY, Yang DD, Samanta Roy DR, Davis RJ, Rakic P, Flavell RA. 1999. The Jnk1 and Jnk2 protein kinases are required for regional specific apoptosis during early brain development. *Neuron* **22**(4): 667-676.

Kuida K, Haydar TF, Kuan CY, Gu Y, Taya C, Karasuyama H, Su MS, Rakic P, Flavell RA. 1998. Reduced apoptosis and cytochrome c-mediated caspase activation in mice lacking caspase 9. *Cell* **94**(3): 325-337.

Kuida K, Zheng TS, Na S, Kuan C, Yang D, Karasuyama H, Rakic P, Flavell RA. 1996. Decreased apoptosis in the brain and premature lethality in CPP32-deficient mice. *Nature* **384**(6607): 368-372.

Lai CS, Fisher SE, Hurst JA, Vargha-Khadem F, Monaco AP. 2001. A forkhead-domain gene is mutated in a severe speech and language disorder. *Nature* **413**(6855): 519-523.

Lander ES Linton LM Birren B Nusbaum C Zody MC Baldwin J Devon K Dewar K Doyle M FitzHugh W et al. 2001. Initial sequencing and analysis of the human genome. *Nature* **409**(6822): 860-921.

Lehto VP. 2001. EGF receptor: which way to go? *FEBS Lett* **491**(1-2): 1-3.

Levine M, Tjian R. 2003. Transcription regulation and animal diversity. *Nature* **424**(6945): 147-151.

Liu FT, Newland AC, Jia L. 2003. Bax conformational change is a crucial step for PUMA-mediated apoptosis in human leukemia. *Biochem Biophys Res Commun* **310**(3): 956-962.

Li WH, Gu Z, Wang H, Nekrutenko A. 2001. Evolutionary analyses of the human genome. *Nature* **409**(6822): 847-849.

Locksley RM, Killeen N, Lenardo MJ. 2001. The TNF and TNF receptor superfamilies: integrating mammalian biology. *Cell* **104**(4): 487-501.

Lossi L, Merighi A. 2003. In vivo cellular and molecular mechanisms of neuronal apoptosis in the mammalian CNS. *Prog Neurobiol* **69**(5): 287-312.

Lowe SW, Lin AW. 2000. Apoptosis in cancer. *Carcinogenesis* **21**(3): 485-495.

Madsen BE, Villesen P, Wiuf C. 2008. Short tandem repeats in human exons: a target for disease mutations. *BMC Genomics* **9**: 410.

Manczak M, Jung Y, Park BS, Partovi D, Reddy PH. 2005. Time-course of mitochondrial gene expressions in mice brains: implications for mitochondrial dysfunction, oxidative damage, and cytochrome c in aging. *J Neurochem* **92**(3): 494-504.

Maksakova IA, Romanish MT, Gagnier L, Dunn CA, van de Lagemaat LN, Mager DL. 2006. Retroviral elements and their hosts: insertional mutagenesis in the mouse germ line. *PLoS Genet* **2**(1): e2.

Marvanova M, Menager J, Bezard E, Bontrop RE, Pradier L, Wong G. 2003. Microarray analysis of nonhuman primates: validation of experimental models in neurological disorders. *FASEB J* **17**(8): 929-931.

McBrearty S, Jablonski NG. 2005. First fossil chimpanzee. *Nature* **437**(7055): 105-108.

McClintock B. 1950. The origin and behavior of mutable loci in maize. *Proc Natl Acad Sci U S A* **36**(6): 344-355.

McClintock B. 1984. The significance of responses of the genome to challenge. *Science* **226**(4676): 792-801.

McClure HM. 1973. Tumors in nonhuman primates: observations during a six-year period in the Yerkes primate center colony. *Am J Phys Anthropol* **38**(2): 425-429.

McConkey EH, Varki A. 2000. A primate genome project deserves high priority. *Science* **289**(5483): 1295-1296.

McDonald JF. 1993. Evolution and consequences of transposable elements. *Curr Opin Genet Dev* **3**(6): 855-864.

McDonnell TJ, Deane N, Platt FM, Nunez G, Jaeger U, McKearn JP, Korsmeyer SJ. 1989. bcl-2-immunoglobulin transgenic mice demonstrate extended B cell survival and follicular lymphoproliferation. *Cell* **57**(1): 79-88.

Medstrand P, van de Lagemaat LN, Dunn CA, Landry JR, Svenback D, Mager DL. 2005. Impact of transposable elements on the evolution of mammalian gene regulation. *Cytogenet Genome Res* **110**(1-4): 342-352.

Meier P, Finch A, Evan G. 2000. Apoptosis in development. *Nature* **407**(6805): 796-801.

Meijerink JP, Mensink EJ, Wang K, Sedlak TW, Sloetjes AW, de Witte T, Waksman G, Korsmeyer SJ. 1998. Hematopoietic malignancies demonstrate loss-of-function mutations of BAX. *Blood* **91**(8): 2991-2997.

Meneveri R, Agresti A, Rocchi M, Marozzi A, Ginelli E. 1995. Analysis of GC-rich repetitive nucleotide sequences in great apes. *J Mol Evol* **40**(4): 405-412.

Mikkelsen T.S. HLW, Eichler E.E., Zody M.C., Jaffe D.B., Yang S.P., Enard W., Hellmann I., Lindblad-Toh K., Altheide T.K., et al.,. 2005. Initial sequencing of the chimpanzee genoem and comparison with the human genome. *Nature* **437**: 69-87.

Miled C, Pontoglio M, Garbay S, Yaniv M, Weitzman JB. 2005. A genomic map of p53 binding sites identifies novel p53 targets involved in an apoptotic network. *Cancer Res* **65**(12): 5096-5104.

Mills RE, Bennett EA, Iskow RC, Luttig CT, Tsui C, Pittard WS, Devine SE. 2006a. Recently mobilized transposons in the human and chimpanzee genomes. *Am J Hum Genet* **78**(4): 671-679.

Mills RE, Luttig CT, Larkins CE, Beauchamp A, Tsui C, Pittard WS, Devine SE. 2006b. An initial map of insertion and deletion (INDEL) variation in the human genome. *Genome Res* **16**(9): 1182-1190.

Mitchell P. 1961. Coupling of phosphorylation to electron and hydrogen transfer by a chemi-osmotic type of mechanism. *Nature* **191**: 144-148.

Morrish TA, Gilbert N, Myers JS, Vincent BJ, Stamato TD, Taccioli GE, Batzer MA, Moran JV. 2002. DNA repair mediated by endonuclease-independent LINE-1 retrotransposition. *Nat Genet* **31**(2): 159-165.

Muzny DM Scherer SE Kaul R Wang J Yu J Sudbrak R Buhay CJ Chen R Cree A Ding Y et al. 2006. The DNA sequence, annotation and analysis of human chromosome 3. *Nature* **440**(7088): 1194-1198.

Needleman SB, Wunsch CD. 1970. A general method applicable to the search for similarities in the amino acid sequence of two proteins. *J Mol Biol* **48**(3): 443-453.

Nesse RM, Bergstrom CT, Ellison PT, Flier JS, Gluckman P, Govindaraju DR, Niethammer D, Omenn GS, Perlman RL, Schwartz MD et al. 2010. Evolution in health and medicine Sackler colloquium: Making evolutionary biology a basic science for medicine. *Proc Natl Acad Sci U S A* **107 Suppl 1**: 1800-1807.

Nijhawan D, Honarpour N, Wang X. 2000. Apoptosis in neural development and disease. *Annu Rev Neurosci* **23**: 73-87.

Normile D. 2001. Genomics. Chimp sequencing crawls forward. *Science* **291**(5512): 2297.

Nutall GHF. 1904. *Blood immunity and blood relationships*. Cambrige University Press, London.

Olson MV, Varki A. 2003. Sequencing the chimpanzee genome: insights into human evolution and disease. *Nat Rev Genet* **4**(1): 20-28.

Ona VO, Li M, Vonsattel JP, Andrews LJ, Khan SQ, Chung WM, Frey AS, Menon AS, Li XJ, Stieg PE et al. 1999. Inhibition of caspase-1 slows disease progression in a mouse model of Huntington's disease. *Nature* **399**(6733): 263-267.

Oomen CA, Girardi CE, Cahyadi R, Verbeek EC, Krugers H, Joëls M, Lucassen PJ. 2009. Opposite effects of early maternal deprivation on neurogenesis in male vs. female rats. *PLoS One* **4** (1): e3675.

Oppenheim RW. 1991. Cell death during development of the nervous system. *Annu Rev Neurosci* **14**: 453-501.

Pardue ML, Lowenhaupt K, Rich A, Nordheim A. 1987. (dC-dA)n.(dG-dT)n sequences have evolutionarily conserved chromosomal locations in Drosophila with implications for roles in chromosome structure and function. *EMBO J* **6**(6): 1781-1789.

Poduri A, Gearing M, Rebeck GW, Mirra SS, Tigges J, Hyman BT. 1994. Apolipoprotein E4 and beta amyloid in senile plaques and cerebral blood vessels of aged rhesus monkeys. *Am J Pathol* **144**(6): 1183-1187.

Polavarapu N, Bowen NJ, McDonald JF. 2006. Identification, characterization and comparative genomics of chimpanzee endogenous retroviruses. *Genome Biol* **7**(6): R51.

Prabhakar S, Noonan JP, Paabo S, Rubin EM. 2006. Accelerated evolution of conserved noncoding sequences in humans. *Science* **314**(5800): 786.

Preuss TM. 2000. What's human about the human brain? In *New cognitive Neurosciences*, (ed. MS Gazzangia), pp. 1219-1234. MIT Press, Cambridge,MA.

Preuss TM, Caceres M, Oldham MC, Geschwind DH. 2004. Human brain evolution: insights from microarrays. *Nat Rev Genet* **5**(11): 850-860.

Puente XS, Velasco G, Gutierrez-Fernandez A, Bertranpetit J, King MC, Lopez-Otin C. 2006. Comparative analysis of cancer genes in the human and chimpanzee genomes. *BMC Genomics* **7**: 15.

Putcha GV, Harris CA, Moulder KL, Easton RM, Thompson CB, Johnson EM, Jr. 2002. Intrinsic and extrinsic pathway signaling during neuronal apoptosis: lessons from the analysis of mutant mice. *J Cell Biol* **157**(3): 441-453.

Raff MC. 1992. Social controls on cell survival and cell death. *Nature* **356**(6368): 397-400.

Rakic S, Zecevic N. 2000. Programmed cell death in the developing human telencephalon. *Eur J Neurosci* **12**(8): 2721-2734.

Rampino N, Yamamoto H, Ionov Y, Li Y, Sawai H, Reed JC, Perucho M. 1997. Somatic frameshift mutations in the BAX gene in colon cancers of the microsatellite mutator phenotype. *Science* **275**(5302): 967-969.

Rasola A, Geuna M. 2001. A flow cytometry assay simultaneously detects independent apoptotic parameters. *Cytometry* **45**(2): 151-157.

Reddy PH, McWeeney S, Park BS, Manczak M, Gutala RV, Partovi D, Jung Y, Yau V, Searles R, Mori M et al. 2004. Gene expression profiles of transcripts in amyloid precursor protein transgenic mice: up-regulation of mitochondrial metabolism and apoptotic genes is an early cellular change in Alzheimer's disease. *Hum Mol Genet* **13**(12): 1225-1240.

Rothenburg S, Eiben M, Koch-Nolte F, Haag F. 2002. Independent integration of rodent identifier (ID) elements into orthologous sites of some RT6 alleles of Rattus norvegicus and Rattus rattus. *J Mol Evol* **55**(3): 251-259.

Rouquier S, Friedman C, Delettre C, van den Engh G, Blancher A, Crouau-Roy B, Trask BJ, Giorgi D. 1998. A gene recently inactivated in human defines a new olfactory receptor family in mammals. *Hum Mol Genet* **7**(9): 1337-1345.

Rutjens E, Balla-Jhagjhoorsingh S, Verschoor E, Bogers W, Koopman G, Heeney J. 2003. Lentivirus infections and mechanisms of disease resistance in chimpanzees. *Front Biosci* **8**: d1134-1145.

Ruvolo M. 1997. Molecular phylogeny of the hominoids: inferences from multiple independent DNA sequence data sets. *Mol Biol Evol* **14**(3): 248-265.

Sabapathy K, Jochum W, Hochedlinger K, Chang L, Karin M, Wagner EF. 1999. Defective neural tube morphogenesis and altered apoptosis in the absence of both JNK1 and JNK2. *Mech Dev* **89**(1-2): 115-124.

Salvesen GS. 2002. Caspases and apoptosis. *Essays Biochem* **38**: 9-19.

Salvesen GS, Dixit VM. 1997. Caspases: intracellular signaling by proteolysis. *Cell* **91**(4): 443-446.

Samonte RV, Eichler EE. 2002. Segmental duplications and the evolution of the primate genome. *Nat Rev Genet* **3**(1): 65-72.

Sarich VM, Wilson AC. 1967. Immunological time scale for hominid evolution. *Science* **158**(805): 1200-1203.

Satta Y, Klein J, Takahata N. 2000. DNA archives and our nearest relative: the trichotomy problem revisited. *Mol Phylogenet Evol* **14**(2): 259-275.

Savage-Rumbaugh ES, Rumbaugh DM, Boysen S. 1978. Symbolic communication between two chimpanzees (Pan troglodytes). *Science* **201**(4356): 641-644.

Savage-Rumbaugh ES, Sevcik RA, Rumbaugh DM, Rubert E. 1985. The capacity of animals to acquire language: do species differences have anything to say to us? *Philos Trans R Soc Lond B Biol Sci* **308**(1135): 177-185.

Scaffidi C, Schmitz I, Krammer PH, Peter ME. 1999. The role of c-FLIP in modulation of CD95-induced apoptosis. *J Biol Chem* **274**(3): 1541-1548.

Seibold H,Wolf RH. 1973. Neoplasms and proliferative lesions in 1065 nonhuman primate necropsies. . *Lab Animal Sci* **23**: 533-539.

Seong GJ, Park C, Kim CY, Hong YJ, So HS, Kim SD, Park R. 2005. Mitomycin-C induces the apoptosis of human Tenon's capsule fibroblast by activation of c-Jun N-terminal kinase 1 and caspase-3 protease. *Invest Ophthalmol Vis Sci* **46**(10): 3545-3552.

Sibley CG, Ahlquist JE. 1987. DNA hybridization evidence of hominoid phylogeny: results from an expanded data set. *J Mol Evol* **26**(1-2): 99-121.

Sinha S, Siggia ED. 2005. Sequence turnover and tandem repeats in cis-regulatory modules in drosophila. *Mol Biol Evol* **22**(4): 874-885.

Slee EA, Adrain C, Martin SJ. 2001. Executioner caspase-3, -6, and -7 perform distinct, non-redundant roles during the demolition phase of apoptosis. *J Biol Chem* **276**(10): 7320-7326.

Snider WD. 1994. Functions of the neurotrophins during nervous system development: what the knockouts are teaching us. *Cell* **77**(5): 627-638.

Sverdlov ED. 2000. Retroviruses and primate evolution. *Bioessays* **22**(2): 161-171.

Syner FN, Goodman M. 1966. Differences in the lactic dehydrogenases of primate brains. *Nature* **209**(5021): 426-428.

Szybalski W, Iyer VN. 1964. Crosslinking of DNA by Enzymatically or Chemically Activated Mitomycins and Porfiromycins, Bifunctionally "Alkylating" Antibiotics. *Fed Proc* **23**: 946-957.

Tait SW, Green DR. 2010. Mitochondria and cell death: outer membrane permeabilization and beyond. *Nat Rev Mol Cell Biol* **11**(9): 621-632.

Tatton WG, Chalmers-Redman R, Brown D, Tatton N. 2003. Apoptosis in Parkinson's disease: signals for neuronal degradation. *Ann Neurol* **53 Suppl 3**: S61-70; discussion S70-62.

Tautz D, Trick M, Dover GA. 1986. Cryptic simplicity in DNA is a major source of genetic variation. *Nature* **322**(6080): 652-656.

Thornberry NA, Rano TA, Peterson EP, Rasper DM, Timkey T, Garcia-Calvo M, Houtzager VM, Nordstrom PA, Roy S, Vaillancourt JP et al. 1997. A combinatorial approach defines specificities of members of the caspase family and granzyme B. Functional relationships established for key mediators of apoptosis. *J Biol Chem* **272**(29): 17907-17911.

Tomasz M. 1995. Mitomycin C: small, fast and deadly (but very selective). *Chem Biol* **2**(9): 575-579.

Tomilin NV. 2008. Regulation of mammalian gene expression by retroelements and non-coding tandem repeats. *Bioessays* **30**(4): 338-348.

Trottier Y, Biancalana V, Mandel JL. 1994. Instability of CAG repeats in Huntington's disease: relation to parental transmission and age of onset. *J Med Genet* **31**(5): 377-382.

Uddin M, Wildman DE, Liu G, Xu W, Johnson RM, Hof PR, Kapatos G, Grossman LI, Goodman M. 2004. Sister grouping of chimpanzees and humans as revealed by genome-wide phylogenetic analysis of brain gene expression profiles. *Proc Natl Acad Sci U S A* **101**(9): 2957-2962.

Vallender EJ, Lahn BT. 2006. A primate-specific acceleration in the evolution of the caspase-dependent apoptosis pathway. *Hum Mol Genet* **15**(20): 3034-3040.

van de Lagemaat LN, Landry JR, Mager DL, Medstrand P. 2003. Transposable elements in mammals promote regulatory variation and diversification of genes with specialized functions. *Trends Genet* **19**(10): 530-536.

van Loo G, Saelens X, van Gurp M, MacFarlane M, Martin SJ, Vandenabeele P. 2002. The role of mitochondrial factors in apoptosis: a Russian roulette with more than one bullet. *Cell Death Differ* **9**(10): 1031-1042.

Vargha-Khadem F, Gadian DG, Copp A, Mishkin M. 2005. FOXP2 and the neuroanatomy of speech and language. *Nat Rev Neurosci* **6**(2): 131-138.

Varki A. 1997. Sialic acids as ligands in recognition phenomena. *FASEB J* **11**(4): 248-255.

Varki A. 2000. A chimpanzee genome project is a biomedical imperative. *Genome Res* **10**(8): 1065-1070.

Varki A, Altheide TK. 2005. Comparing the human and chimpanzee genomes: searching for needles in a haystack. *Genome Res* **15**(12): 1746-1758.

Varki A,Nelson DL. 2007. Genomic comparisons of humans and chimpanzees. *Annu Rev Anthropol* **36**: 191-209.

Vaux DL, Cory S, Adams JM. 1988. Bcl-2 gene promotes haemopoietic cell survival and cooperates with c-myc to immortalize pre-B cells. *Nature* **335**(6189): 440-442.

Venter JC Adams MD Myers EW Li PW Mural RJ Sutton GG Smith HO Yandell M Evans CA Holt RA et al. 2001. The sequence of the human genome. *Science* **291**(5507): 1304-1351.

Verkerk AJ, Pieretti M, Sutcliffe JS, Fu YH, Kuhl DP, Pizzuti A, Reiner O, Richards S, Victoria MF, Zhang FP et al. 1991. Identification of a gene (FMR-1) containing a CGG repeat coincident with a breakpoint cluster region exhibiting length variation in fragile X syndrome. *Cell* **65**(5): 905-914.

Verstrepen KJ, Reynolds TB, Fink GR. 2004. Origins of variation in the fungal cell surface. *Nat Rev Microbiol* **2**(7): 533-540.

Volfovsky N, Oleksyk TK, Cruz KC, Truelove AL, Stephens RM, Smith MW. 2009. Genome and gene alterations by insertions and deletions in the evolution of human and chimpanzee chromosome 22. *BMC Genomics* **10**: 51.

Walker LCaC, L.C. 1999. *Alzheimer's Disease*. Lippincott Williams and Wilkins, Philidelphia.

Wallace-Brodeur RR, Lowe SW. 1999. Clinical implications of p53 mutations. *Cell Mol Life Sci* **55**(1): 64-75.

Wang HG, Millan JA, Cox AD, Der CJ, Rapp UR, Beck T, Zha H, Reed JC. 1995. R-Ras promotes apoptosis caused by growth factor deprivation via a Bcl-2 suppressible mechanism. *J Cell Biol* **129**(4): 1103-1114.

Wang QF, Prabhakar S, Chanan S, Cheng JF, Rubin EM, Boffelli D. 2007. Detection of weakly conserved ancestral mammalian regulatory sequences by primate comparisons. *Genome Biol* **8**(1): R1.

Watanabe H, Fujiyama A, Hattori M, Taylor TD, Toyoda A, Kuroki Y, Noguchi H, BenKahla A, Lehrach H, Sudbrak R et al. 2004. DNA sequence and comparative analysis of chimpanzee chromosome 22. *Nature* **429**(6990): 382-388.

Waters DJ, Sakr WA, Hayden DW, Lang CM, McKinney L, Murphy GP, Radinsky R, Ramoner R, Richardson RC, Tindall DJ. 1998. Workgroup 4: spontaneous prostate carcinoma in dogs and nonhuman primates. *Prostate* **36**(1): 64-67.

Whiten A, Goodall J, McGrew WC, Nishida T, Reynolds V, Sugiyama Y, Tutin CE, Wrangham RW, Boesch C. 1999. Cultures in chimpanzees. *Nature* **399**(6737): 682-685.

Williams GC. 1957. Pleiotropy, Natural Selection and the Evolution of Senscence. *Evolution* **11** (4): 398-411.

Williams MF. 2002. Primate encephalization and intelligence. *Med Hypotheses* **58**(4): 284-290.

Winter H, Langbein L, Krawczak M, Cooper DN, Jave-Suarez LF, Rogers MA, Praetzel S, Heidt PJ, Schweizer J. 2001. Human type I hair keratin pseudogene phihHaA has functional orthologs in the chimpanzee and gorilla: evidence for recent inactivation of the human gene after the Pan-Homo divergence. *Hum Genet* **108**(1): 37-42.

Wood B, Collard M. 1999. The human genus. *Science* **284**(5411): 65-71.

Wu Q. 2005. Comparative genomics and diversifying selection of the clustered vertebrate protocadherin genes. *Genetics* **169**(4): 2179-2188.

Wyllie AH. 1980. Glucocorticoid-induced thymocyte apoptosis is associated with endogenous endonuclease activation. *Nature* **284**(5756): 555-556.

Yamakoshi G, Myowa-Yamakoshi M. 2004. New observations of ant-dipping techniques in wild chimpanzees at Bossou, Guinea. *Primates* **45**(1): 25-32.

Yamada T, Ohtani S, Sakurai T, Tsuji T, Kunieda T, Yanagisawa M. 2006. Reduced expression of the endothelin receptor type B gene in piebald mice caused by insertion of a retroposon-like element in intron 1. *J Biol Chem* **281**(16): 10799-10807.

Yan H, Yuan W, Velculescu VE, Vogelstein B, Kinzler KW. 2002. Allelic variation in human gene expression. *Science* **297**(5584): 1143.

Yang J, Liu X, Bhalla K, Kim CN, Ibrado AM, Cai J, Peng TI, Jones DP, Wang X. 1997. Prevention of apoptosis by Bcl-2: release of cytochrome c from mitochondria blocked. *Science* **275**(5303): 1129-1132.

Yee HA, Wong AK, van de Sande JH, Rattner JB. 1991. Identification of novel single-stranded d(TC)n binding proteins in several mammalian species. *Nucleic Acids Res* **19**(4): 949-953.

Yohn CT, Jiang Z, McGrath SD, Hayden KE, Khaitovich P, Johnson ME, Eichler MY, McPherson JD, Zhao S, Paabo S et al. 2005. Lineage-specific expansions of retroviral insertions within the genomes of African great apes but not humans and orangutans. *PLoS Biol* **3**(4): e110.

Yoshida H, Kong YY, Yoshida R, Elia AJ, Hakem A, Hakem R, Penninger JM, Mak TW. 1998. Apaf1 is required for mitochondrial pathways of apoptosis and brain development. *Cell* **94**(6): 739-750.

Yunis JJ, Prakash O. 1982. The origin of man: a chromosomal pictorial legacy. *Science* **215**(4539): 1525-1530.

Yunis JJ, Sawyer JR, Dunham K. 1980. The striking resemblance of high-resolution G-banded chromosomes of man and chimpanzee. *Science* **208**(4448): 1145-1148.

Zhang J. 2003. Evolution of the human ASPM gene, a major determinant of brain size. *Genetics* **165**(4): 2063-2070.

Zhang J, Webb DM, Podlaha O. 2002. Accelerated protein evolution and origins of human-specific features: Foxp2 as an example. *Genetics* **162**(4): 1825-1835.

Zhang XM, Cathala G, Soua Z, Lefranc MP, Huck S. 1996. The human T-cell receptor gamma variable pseudogene V10 is a distinctive marker of human speciation. *Immunogenetics* **43**(4): 196-203.

Ziegler U, Groscurth P. 2004. Morphological features of cell death. *News Physiol Sci* **19**: 124-128.